DTU Health Tech
Department of Health Technology

Pupil behavior in listening and speaking time of interactive communication

Pioneering exploration of cognitive demands in effortful conversations

Master Thesis

Pupil behavior in listening and speaking time of interactive communication
Pioneering exploration of cognitive demands in effortful conversations

Master Thesis
July 2, 2023

By
Pablo Castro Ilundain

Cover photo: Eriksholm Research Centre, 2020
Published by: DTU, Department of Health Technology, Ørsteds Plads, Building 352, 2800 Kgs. Lyngby Denmark
www.hea.healthtech.dtu.dk
ISSN: [0000-0000] (electronic version)
ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)
ISBN: [000-00-0000-000-0] (printed version)

Abstract

This MSc thesis conducts an analysis of pupil behavior and eye gaze movements during interactive communication complementing the ongoing AMEND (v**A**lid outco**M**e m**E**asure for commu**N**ication **D**ifficulty) [1] research project. This thesis was made possible thanks to a collaboration with the company Eriksholm Research Centre, which is part of Oticon. Pupillometry and eye gaze tracking data are collected within specific speaking and listening windows, which are further processed and analyzed in order extract meaningful conclusions. The study examines the cognitive load or listening effort experienced during interactive communication, while also exploring the influence of various noise conditions and hearing aid settings. As this is a pioneering project in the field, this research lays the groundwork for what should be the expected outcomes on any similar work within the field. Giving out insights into the understanding of communication dynamics as well as contributing to quantifying the impact of various conditions and hearing aid settings on communication processes.

Acknowledgements

I would like to thank greatly both supervisors of this thesis, as they've always been there to help and guide me in the right direction. This whole Msc thesis would not have been possible without your knowledge, support and mentoring. Thank you so much to both:

- **Susan Aliakbary Hosseinabadi** (sulb@eriksholm.com), Scientist, Eriksholm Research Centre.
- **Dorothea Wendt** (dowe@eriksholm.com), Principal Scientist, Eriksholm Research Centre.

I would also like to thank all the colleagues from Eriksholm Research Centre for the unconditional support, motivation and gratitude that they have given me all throughout my master thesis research as well as some of the other projects related to my 'Student Assistant' position.

I also want to thank my partner, Helen, not only for being supportive every step of the way but also for being my main source of motivation and happiness from the first day I arrived to Denmark until today.

Lastly, I want to thank DTU itself for giving me the opportunity of a lifetime and I specially want to thank all the friends I've made thanks to DTU, which come from different courses, clubs and even departments.

Contents

Nomenclature and Abbreviations

The next list describes several abbreviations and symbols that will be later used within the body of the document

AMEND	v**A**lid outco**M**e m**E**asure for commu**N**ication **D**ifficulty
ANS	Autonomous Nervous System
CSC	Communicative State Classification
DSST	Digit Symbol Substitution Test
EEG	Electroencephalogram
FTO	Floor Transfer Offset
HINT	Hearing In Noise Test
HI	Hearing Impaired
HL	Hearing Level
IPU	Inter-Pausal Unit
LMM	Linear Mixed Model
MAD	Median Absolute Deviation
MoCA	Montreal Cognitive Assessment
MPD	Mean Pupil Dilation
NAN	Not A Number
NH	Normal Hearing
PNS	Parasympathetic Nervous System
PPD	Peak Pupil Dilation
PTA	Pure Tone Audiometry
REM	Real Ear Measurement
RMS	Root Mean Square
RT60	Reverberation Time
SHL	Simulated Hearing Loss
SNR	Signal-to-Noise Ratio
SNS	Sympathetic Nervous System
SPL	Sound Pressure Level
TP	Test Participant
VAD	Voice Activity Detection
VBDS	Visual Backward Digit Span

1 Introduction

1.1 Context and motivation

Communication is a dynamic and complex process as it needs to consider the behaviour from the interlocutor(s). Many researches have been conducted to investigate details of communication by using linguistic, social, psychological and anthropological approaches. Difficulty in communication is one of the most disabling consequences for people living with hearing loss. Individuals that suffer from any type of hearing loss will most likely find it challenging to engage, follow and comprehend daily conversations, specially in noisy conditions, which usually results in an increment of difficulty in regular day-to-day situations. In the long run, communication difficulties ultimately lead to social isolation, reduced self-esteem and increased stress and anxiety.

As described by Andersen et al. in 1982 [2], presbycusis is a very common slowly progressive impairment of the auditory pathways that starts during the fourth decade of life that results in gradual hearing losses. This decline is more pronounced for higher frequencies compared to lower frequencies. Research has given evidence of a notable correlation between hearing loss and challenges in communication [3], meaning that speech intelligibility is also affected greatly by hearing loss. This has lead to the fact that more recent studies aimed to study the effect of communication difficulty on various groups of communication outcomes.

Knowing this, a handful of studies suggest that physiological measures such as pupillometry [4] [5] [6], heart rate [7] and electroencephalography (EEG) [8] contain important information about different cognitive processes like attention, perception and sensation. This processes are mainly controlled by the autonomous nervous system (ANS). This system is divided in distinct components that are antagonistic to each other, these are the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). These mechanisms are responsible for a ”fight-or-flight” or a ”feed-and-breed” response, respectively. Any sudden change in a physiological measure can then be explained as a body response from the sympathetic nervous system to some stimuli.

The main objective of this project, that this master’s thesis is associated with, is to estimate or predict a global marker or indicator that resembles communication difficulty during interactive communication by using different communication outcomes. Furthermore, this thesis will primarily concentrate on the examination of pupillometry and fixation duration data. These metrics measure both the dilation (expansion) or contraction of the pupils and the duration of gaze fixation on a specific point over a period of time, respectively. The main focus is on the processed responses of multiple test participants (normal hearing and hearing impaired) in different conditions. These measures, when analyzed properly, could give an insight into the mechanisms underlying listening or communication difficulties specially for population that suffer from hearing loss. On the long run, this could end up contributing to the development of more advanced and effective communication technologies to be implemented in assistive devices such as hearing aids.

As a matter of fact, the World Health Organization (WHO) [9] has reported that by 2050, nearly 2.5 billion people will live with some degree of hearing loss and at least 700 million of those people will require rehabilitation services such as hearing aids. The scientific community, as well as hearing aids manufacturers have invested a lot of resources and effort to investigate different strategies, not only for the assessment and treatment of hearing impairments, but also for the development of new and improved devices. Additionally, there is great interest on developing optimal strategies for fitting the hearing aids. The availability, distribution and overall quality of hearing aids have come a long way but these devices still face some limitations.

One of the biggest challenges that individuals with hearing loss can normally come across on their day to day life is the ”cocktail party problem”, which happens whenever there’s a considerable amount of background noise somewhere caused by many different individuals talking simultaneously. This situation makes it almost impossible for the hearing impaired to focus attention or to differentiate any individual voices. This, and many other hearing challenges could easily be tackled with future technology by researching on selective attention, listening and/or speaking effort and communication difficulty.

1.2 Background

This thesis is classified within the research category of cognitive studies, specifically exploring interactive communication. It utilizes pupillometry, eye gaze derived measures and features to examine cognitive aspects of different phases of the communication process (speaking and listening). Research projects within this category strive to discover, unveil and deliver a deeper understanding or knowledge about the human auditory system, communication dynamics and cognitive processes. This type of projects usually involve many different test participants, a controlled lab environment with precise measurement devices and/or complex modelling to obtain data that, after a processing and filtering phase, can draw meaningful conclusions that could potentially lead up to the discovery of cognitive behavioral processes. Although certain aspects can be controlled within an experiment setup, external factors like light conditions or subjects’ limitations may introduce some variability. This pioneer research project is unique, as no previous research to this day have explored pupillometry or eye movement behavior in an interactive communication setting, more specifically, within communication dynamics of listening and speaking.

As mentioned in [4], psychologists have been involved with pupillary studies since the early 1960s when many different research labs studying experimental psychology and psychopathology published various reports. The most controversial approaches at the time were presented on a series of papers written both by Eckhard H. Hess and James M. Polt: *”Pupil Size as Related to Interest Value of Visual Stimuli”* [10] and *”Pupil Size in Relation to Mental Activity during Simple Problem Solving”* [11]. These papers stated that pupil dilation occurs in response to positive affect stimuli and constriction in response to negative affect. Some years later, Hess and Polt also explored pupil dilation during mental activities, which was later on picked up by other scientists, such as Kahneman and Beatty, who would work on the development of some of the concepts of cognitive psychology.

Linguistics is a interdisciplinary field of study that investigates many different aspects of human language (some examples can be seen in table 1.1). In 1974, Harvey Sacks, Emanuel Schegloff and Gail Jefferson developed a method for conversation analysis in the paper *”A Simplest Systematics for the Organization of Turn-Taking for Conversation”* [12], which consists on examining the structure and organization of speech in interactive conversations. The influence of this method is unmatched, as it is still used to this day as a basis for the classification and/or definition of conversational features such as gaps, pauses, turns and overlaps. These terms, when combined, create the frame from which a conversation is built upon. Intrinsic characteristics such as the number of repetitions, order of appearance, durations and so on, will portray lots of information about communication difficulty, conversational effort and many more.

Table 1.1: Basic aspects or categories of linguistics and their respective focus areas.

Categories	Focus area
Phonetics	Sounds in language
Phonology	Organization of sounds in language
Morphology	Structure of words
Syntax	Structure of sentences
Semantics	Meaning in language
Pragmatics	Use of language in context

1.3 Hypothesis

The purpose of this thesis is to gain a more profound understanding of communication effort, investment and arousal in a realistic conversation by analyzing pupil diameter and fixation duration measures in different conditions, both for individuals with age-related normal hearing (NH) as well as hearing impaired (HI). The key motivation is to extract and examine different features from pupillometry data categorized in different experiment conditions to check whether or not there is a clear connection between them.

As previously stated, this thesis is a research collaboration with Eriksholm Research Centre, or more specifically the AMEND project [1]. In terms of timeline, the AMEND project planned for two years with two experimental phases: AMEND I, which involved dyadic conversation between age-related NH participants with completed data collection in Spring 2022. AMEND II conducted conversation between NH-HI participants which data collection done in Spring 2023.

Most conventional methods of hearing rehabilitation don't really take into account an individual's practical ability to communicate, understand and/or process any conversation. This is mainly due to the nature of human communication, which makes it almost impossible to objectively quantify the difficulty or effort from the individuals involved. For this reason, many recent studies such as the one from Haro et. al. (2021) [13], have started to correlate physiological measures to the aforementioned conversation difficulty just to grasp into how these mechanisms work and what can be predicted to happen in different scenarios. This knowledge can not only be used in the development of advanced rehabilitation devices for the cognitive impaired individuals, but also in the initial assessment and/or intervention strategies so that the whole extent of a person's impairment is known and treated properly.

This thesis tested the validity of the following hypotheses:

- **H1** : Cognitive load will rise with increasing task demands during speaking and listening, which will be reflected on the response from both pupil size and fixation duration measures.
- **H2** : Pupillometry features can serve as markers for cognitive load or effort in both cohorts of normal hearing and hearing impaired participants.
- **H3** : Pupil size and fixation duration responses during speaking and listening time windows will exhibit variations across different noise conditions, indicating a correlation with communication difficulty or effort. It is expected that higher levels of task

demand elicit equally higher responses.

- **H4** : The effectiveness of hearing aids (with and without noise suppression) will be reflected when comparing aided and unaided results across different noise conditions.

1.4 Goals and methods

As previously mentioned, both AMEND I and AMEND II are pioneer projects in the sense that there are no other examples of pupillometry or eye gaze related studies involving interactive communication between two participants which have to solve a diapix [14] task. Based on this, the results that come out of this study are quite important to the scientific community as they will help set the bar on what type of behavioral pupil responses should be expected, specially those regarding listening and speaking times in a conversation. Hypothesis will also be tested out and results will be compared to many other similar pupillometry studies to convey similarities in attention and effort decoding and whether if there's any similar behavioral pattern happening in a conversation in comparison to just listening to a stimulus.

This research also aims to provide a deeper understanding on how cognitive processes happen in real-time and how can they be translated into more meaningful principles such as attention, effort, memory and so on. Mainly pupillometry and fixation duration responses and/or features from different speaking or listening time periods are going to be used to accomplish the aforementioned research goal.

2 Methods

The aim of this chapter is to lay the theoretical foundations and concepts that will serve as the basis for the interpretation and analysis of the results presented in the following chapters. Additionally, this chapter will also provide information regarding the experimental setup, including the participants involved, the workflow of the tasks, as well as the different noise conditions and hearing aid settings.

2.1 Test setup and apparatus

2.1.1 Sound Wave Lab

The Sound Wave lab at Eriksholm Research Centre was used throughout the whole project to gather data for every Test Participant (TP). A schematic of the room (4.35 x 3.57 x 2.5 [m]) can be seen in fig. 2.1, the room was equipped with:

- Eight loudspeakers (Genelec Z8000a; Lisalmi, Finland) spread equally distanced in a ring with diameter of 2.5 [m] .
- Two Tobii 3 glasses (see section 2.1.2 for more information).
- Eight cameras (Vicon Vero; Oxford, UK) placed on the walls that form a motion tracking system.
- Two wireless head-worn microphones (DPA 4488; Allerød, Denmark).
- Two wristbands (Empatica E4; Milano, Italy).
- One sound card (FerroFish Pulse 16MX; Linz am Rhein, Germany) connected to a computer running MATLAB [15] (Matlab R2021a; Natick, USA).
- Three trigger-boxes to help synchronize all simultaneous recordings.

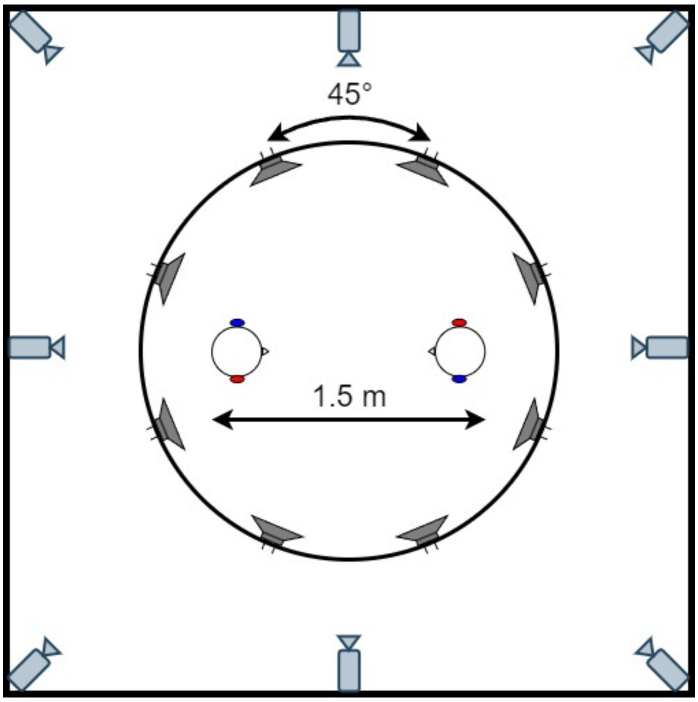


Figure 2.1: Diagram of the test room (Sound Wave Lab, Eriksholm Research Centre).

MATLAB [15] was used in order to start each of the trials and to synchronize audio inputs/outputs as well as the trigger-boxes. Data collection from other devices was done by using their associated proprietary software such as Tobii Glasses 3 Controller, Vicon Nexus and Empatica E4 Realtime. The room can be classified as non-reverberant, as indicated by the measured reverberation time (RT_{30}) of around 0.23 seconds.

2.1.2 Eye-tracking device

Eye-tracking data were recorded continuously for each test participant across entire experiment. A pair of wearable eye trackers were used for this purpose (Tobii Pro Glasses 3; Danderyd, Sweden), as seen in fig. 2.2. In AMEND II, if any TP had any type of vision impairment such as farsightedness (hyperopia) or nearsightedness (myopia), corrective lenses were added to the glasses.

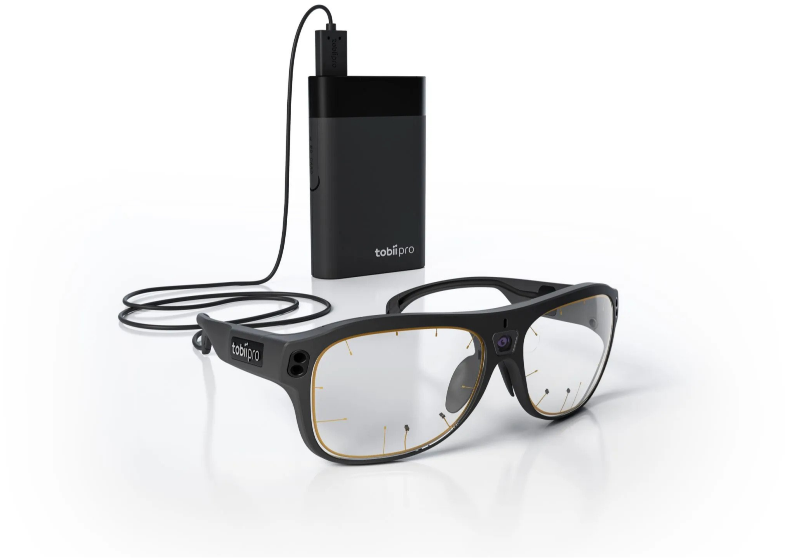


Figure 2.2: Tobii Pro Glasses 3: Head and recording units.

Data was collected at a sampling frequency of 50 Hz, capturing pupil size or diameter (in mm) for each pupil, the relative 2D and 3D gaze (in mm), and the corresponding sample validity of these measurements. All measurements were conducted in the same lab with almost identical luminosity levels across visits.

2.2 Test participants

As previously stated in section 1.3, each of the parts of AMEND were carried out by different test participants:

- **AMEND I:** 24 age-related NH participants (13 females, 11 males) with age range of 56 to 73 years old. Average hearing thresholds were 26 dB Hearing Level (HL) across six octave frequency bands from 250 Hz to 8 kHz.
- **AMEND II:** 12 age-related NH participants and 12 HI participants with different severity levels of hearing loss ranging from 30 to 65 dB HL. Their age range was from 51 to 72 years old. Average hearing thresholds were 25 dB HL for the NH participants and 43 dB HL for the HI participants.

All participants enrolled in the project were native danish speakers with an acceptable eyesight that permitted them, with or without correction lenses, to see clearly up to a distance of 1.5 m. Homogeneous groups of pairs with mixed gender and hearing were created for the experimental phase (AMEND I: NH pairs, AMEND II: NH paired with HI). To ensure impartiality and prevent any biases or interpersonal communication strategies, all recruited participants were mutually unfamiliar with one another. Lastly, participants were requested to sign a consent form approved by the Science-Ethics Committee for the Capital Region of Denmark (reference number H-16036391) and they were paid a small monetary compensation for their participation.

2.3 Test workflow

In both AMEND I and AMEND II, NH participants had a single visit to Eriksholm where their hearing was initially assessed through an audiometry, followed by the activities from visit 2. However, for HI participants in AMEND II, the activities were divided into two separate visits to Eriksholm.

2.3.1 Visit 1

During the first visit, participants underwent several cognitive and auditory assessments in order to estimate how challenging will the next visit be and also to provide a baseline of how each participant will perform, based on their estimated capacities or abilities. The estimated time this visit takes is around 2 hours for HI and 1 hour for NH participants.

Firstly, the cognitive tests were: the Montreal Cognitive Assessment (MoCA) [16], which was used to screen or detect mild cognitive impairment. The Digit Symbol Substitution Test (DSST) [17] was administered to evaluate cognitive speed, attention, and working memory. Similarly, the Visual Backward Digit Span (VBDS) [18] test was used to measure working memory capacity and cognitive strategies used to solve the task.

On the other hand, the audiometry-related assessments were: the Hearing in Noise Test (HINT) [19], which was used to assess speech intelligibility in noisy environments, while Real Ear Measurement (REM) and Pure Tone Audiometry (PTA) were used to evaluate hearing sensitivity and the amplification effect of hearing aids inside the ear.

2.3.2 Visit 2

When participants arrived to the facilities, they were greeted by a qualified clinical audiologist who first explained the experimental procedure and the task itself, then asked the TPs to sign their consent and, lastly, carry out the necessary audiometry measurement (which will also be compared to the one realized in Visit 1). Then, the participants were taken to the lab and they were seated face to face, as stated in fig. 2.1. The experimenters then fitted the participants the Tobii glasses and the head-worn microphones (as well as some other apparatus not related to the contents of this thesis).

In order for participants to engage in a face-to-face dyadic conversation, the subjects were asked to solve a Diapix task [14] by looking for a maximum of 12 differences between two different drawings (one for each TP). These drawings depict 3 possible scenarios, a beach, a forest or a city (ordered in trios, maintained from trial to trial) and they belong to the original Diapix corpus (Baker & Hazan, 2011) [14] but with the original text translated to danish. Once the participants were familiarized with the task, they went through a training phase which consisted on them practicing the Diapix task for around 30 seconds in the ”Quiet” condition, to ensure that they were able to make a balanced conversation during the test.

Measurements were then carried out with the participants working together to solve the task. The time limit for each trial was set to be 4 minutes and if the participants found

12 differences before this time, the test was halted and the completion time noted. The estimated time it takes to fully complete this visit was around 2.5 hours.

2.4 Test conditions

As seen previously in section 2.3.2, the second visit involved the main experiment trials, realized with different conditions. These conditions differ from one phase of the project to another, as mentioned in section 1.3:

- **AMEND I:** Four test conditions: Quiet, SHL (Simulated Hearing Loss), N60 and N70. In both the Quiet and SHL conditions, there was no background noise (N0). However, in the Quiet condition, the participants had non-occluded ear canals, while in the SHL condition, their ear canals were occluded with a pair of earplugs (Alpine, MusicSafe Pro; Soesterberg, Netherlands), which were used for simulating a (conductive) hearing loss of around 25 dB in average across the regular octave frequency bands. For the other two conditions, multi-talker babble noise was presented at 60 (N60) and 70 (N70) dBA from all the loudspeakers, respectively. The noise level was calibrated with a sound level meter (Bruel & Kjær 2250; Nærum, Denmark) and a reference microphone placed in both talkers positions.
- **AMEND II:** The noise conditions remained identical to those in AMEND I, with the exception of the SHL condition, which was eliminated due to the impracticality of completion for HI participants. The three noise conditions, Quiet, N60 and N70 were paired with different scenarios for the HI participants. In the Aided condition (AA or AB), HI participants wore hearing aids with different settings (A or B), while in the unaided condition (UN), they did not wear hearing aids at all. The hearing aids (Oticon More 1; Smørum, Denmark) were fitted using VAC+ (Oticon standard) for each HI participant and the dome type was chosen by the support audiologist, which was mainly the one that HI participants were more familiar with (open or power domes preferred more than closed domes). Settings A or B correspond to the hearing aids Help Level (HL) of 3 and 6. Sanchez Lopez [20] demonstrated, and as it can be seen both in table A.1 and in fig. A.1, that the primary distinction between settings HL3 (AA) and HL6 (AB) lies in the amount of noise reduction they apply, leading to big contrasts in speech intelligibility. Finally, the combinations of N0, N60, N70 with UN, AA, AB yield a total of 9 different test conditions.

2.5 Data analysis and processing

From all the data gathered by all different apparatus mentioned in section 2.1.1, only speech and eye tracking data will be further analyzed, as they form the basis of this thesis. This section will evaluate gathered data from all trials both from AMEND I and AMEND II, which were automatically split in different audio or eye tracking files that were synchronized with each other (trial by trial) thanks to the trigger boxes. It is also important to mention that both eye tracking and speech data were gathered and also processed using MATLAB [15] (Matlab R2021a; Natick, USA).

2.5.1 Preprocessing

The Tobii Pro Glasses 3, as described in section 2.1.2, is a device that captures many different eye features simultaneously. This allowed for the investigation or exploration of other eye characteristics (such as gaze) rather than just pupillometry, with the intention of testing or identifying whether if any of these features is an indicator of actual cognitive load for the TP.

Preprocessing is an essential step or phase in eye tracking data analysis that, given the raw data from the measurements, removes outliers and/or artifacts that are expected to happen due to some of the factors listed in section 2.5.2. The purpose of the preprocessing phase is to ensure the quality and consistency of the data that will be subsequently used for the analysis. It is common practice in pupillometry and eye gaze related studies to choose one preprocessing method or pipeline but, in this thesis, all data including pupil size and eye gaze will be preprocessed with two different methods (described in section 2.5.1 and section 2.5.1), which will serve as a robust form of validation between both methods. This approach allows a comprehensive assessment and comparison of the performance and reliability of both methods.

Pupil data preprocessing

According to the literature [21], there isn't a standard preprocessing method for pupil data as it should be chosen according to the type of experiment, the setup/capture device chosen, its sampling frequency and also, on the duration of the measurements.

In this thesis, the pre-processing of the pupillometry primarily focused on the removal of outliers, following the method described by Kret et al. (2018) [22]. This method aims to identify three different types of invalid/erroneous pupil size samples (which can be seen in fig. 2.3): **(1)** *Dilation speed outliers and edge artifacts:* Samples characterized by having values that significantly differ from their adjacent samples; **(2)** *Trend-line deviation outliers:* samples with a disproportionate deviation from a smooth pupil trend line; **(3)** *Temporally isolated samples:* bogus samples separated in time in small chunks, easily identifiable.

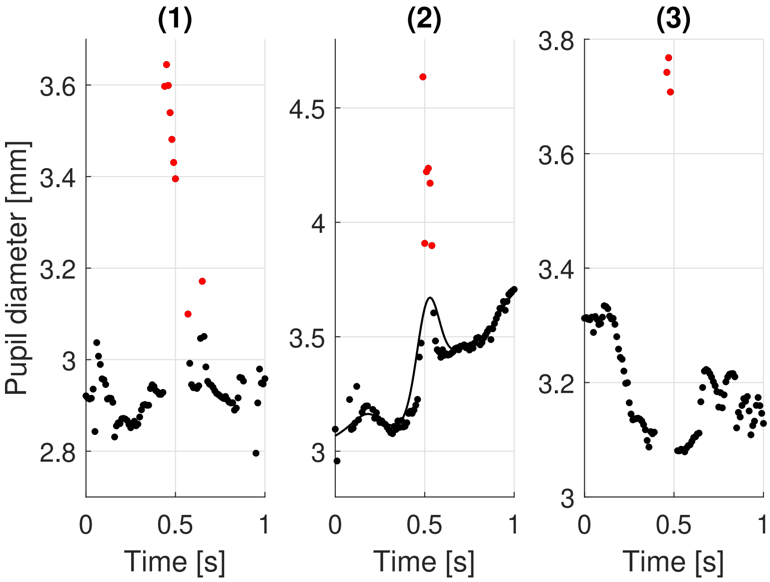


Figure 2.3: Raw pupil size data showing the three aforementioned types of erroneous samples (**(1)**, **(2)** and **(3)**) that will be the target for the preprocessing pipeline defined by Kret et al. (2018).

Due to the existence of gaps or non-uniform sampling in the data, the changes in value from sample to sample cannot be compared directly. This issue is solved by Kret et al. (2018) [22] with the introduction of the concept of *normalized dilation speed*, which is the ratio between the absolute change of the values in samples and its separation in time.

Being ($d(i)$) the pupil dilation time series with timestamps ($t(i)$), the normalized dilation speed ($d'(i)$) at each sample (i), can be calculated as the maximum absolute normalized change relative to either the prior or subsequent sample,

$$d'(i) = max\left(\left|\frac{d(i) - d(i-1)}{t(i) - t(i-1)}\right|, \left|\frac{d(i+1) - d(i)}{t(i+1) - t(i)}\right|\right). \tag{2.1}$$

To detect dilation speed outliers, the median absolute deviation (MAD) method is utilized, which was defined by Leys et al. (2013) [23]. The MAD method is a reliable dispersion metric that is not affected by outliers and it is calculated as,

$$\text{MAD} = median\left(\left|d'(i) - median(d'(i))\right|\right). \tag{2.2}$$

By using a constant (n), the threshold from which the dilation speed outliers are defined can be calculated as,

$$\text{Threshold} = median(d'(i)) + n \cdot \text{MAD}. \tag{2.3}$$

Any sample with a calculated dilation speed above the threshold is now defined as an outlier and will be rejected. Also, in order to remove edge artifacts that may occur around gaps, typically caused by eye blinks, samples within 50 ms of gaps will be rejected. In this case, a gap is defined as any data sections without values (designated as ”Not A Number” (NaN)) with a duration of at least 75 ms.

Samples that have a strong deviation from the trend line, which is generated right after the dilation speed threshold is applied, will also be rejected. A new trend line can be generated, using the data without the newly rejected samples, and it is possible that some of the previously discarded samples are reintroduced because their deviation from the new trend line is within range. A sparsity filter is also introduced by Kret et al. (2018) [22] to remove samples bordering gaps smaller than 50 ms, this ensures that noisy samples as well as those generated by a glitch in the eyetracker are removed from the data.

Gaze data preprocessing
Gaze data was preprocessed using the ”PUPILS” pipeline, which was released together with an article created by Relaño-Iborra et al. (2020) [24]. The pipeline classifies pupillometry samples as three possible events such as blinks, saccades and fixations and then, preprocesses the raw pupil size data accordingly. In our case, ”PUPILS” is used only for detecting fixations in the data, as those time chunks will later be used for calculating fixation duration.

In order for the pipeline [24] to detect blinks, an alternative method was used instead of utilizing the *normalized dilation speed* as discussed earlier in section 2.5.1. This method states that, a blink is defined as region of outlier samples where the measured pupil size is smaller than three standard deviations below the average pupil size of the entire trial, as it is recommended by Winn et al. (2018) [21]. Additionally, data chunks containing ”NaN” or ”0” values were also classified as blinks.

The ”PUPILS” pipeline [24] can either calculate the angular velocity for X and Y or it can also accept the angular velocity as an input parameter in order to use it for the detection of saccades. Tobii Pro Glasses 3 estimates three-dimensional (3D) gaze with a specific coordinate system in which the axes are oriented as follows: when viewed from the perspective of the subject, the X-axis points horizontally to the left side, the Y-axis points vertically upwards and the Z-axis points forward. Being the origin (0,0,0), the middle point between the two eyes placed within the glasses. Therefore, the angular velocity, v_{gaze}, was given to ”PUPILS” as an input, and it was calculated as:

$$v_{gaze} = \sqrt{v_\theta^2 * cos(\phi)^2 + v_\phi^2}, \tag{2.4}$$

being θ the azimuth, ϕ the elevation and also v_θ and v_ϕ their respective velocities, which were calculated as,

$$v_\theta = \frac{\Delta\theta}{\Delta t} \quad \text{and} \quad v_\phi = \frac{\Delta\phi}{\Delta t}. \tag{2.5}$$

Both the azimuth, θ, and the elevation, ϕ, are the result of the transformation of 3d gaze samples from cartesian to spherical coordinates. A saccade is detected whenever the angular velocity, v_{gaze}, exceeds a certain velocity threshold, which in this case is set to 30 [degrees/s] or [°/s]. Also, in order to classify an event as a saccade, its duration must be of at least 10 ms.

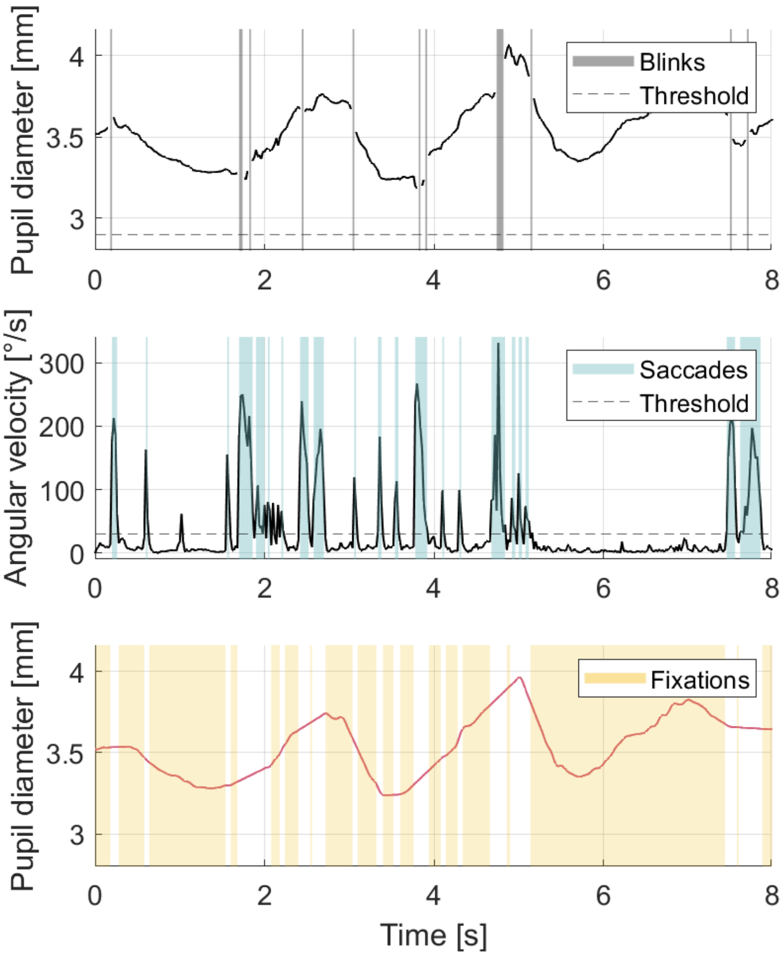


Figure 2.4: Event detection or classification steps in the ”PUPILS” pipeline by Relaño-Iborra et al. (2020) [24], ordered from top to bottom (blink, saccade and fixation events). The events are presented along with the corresponding data (pupil diameter or angular velocity) that contributed to their estimation. In this example, blink events were detected because of gaps in the data (NaNs), not because of the standard deviation threshold.

Once blinks and saccades are already classified, fixations are defined as all the remaining time intervals in the data which haven't been classified yet. This fixation events will be the time chunks from which samples will be used in order to calculate the fixation duration for each trial. An example of the sequence of classification of events can be seen in fig. 2.4.

2.5.2 Data inclusion requirements

As it is mandatory on any study involving eye-tracking data, it is always important to exclude trials that don't comply with certain criteria aimed at ensuring some degree of accuracy and integrity in the data.

Table 2.1: Total number of trials in AMEND I and AMEND II, from the descriptions seen in section 2.2 and section 2.4.

	Conditions	Repetitions	TPs	Total trials	Non-rejected trials
AMEND I	4	2	24	192	**131**
AMEND II	9	1	24	216	**123**

The total number of trials per part of the project, as seen in table 2.1, is not equivalent to the actual number of files selected for data analysis due to factors such as:

- **Missing recordings:** Due to data not recorded in one or both devices (microphones and/or glasses), which renders the analysis impossible. Or due to connectivity problems, storage or battery issues, etc.
- **Un-calibrated eye-tracking data:** Whenever the eye-tracking calibration, was too difficult to perform for any of the TPs, it was skipped and this resulted in data loss, gaze estimation inaccuracies and an overall poor data quality.
- **Individual eye disorders:** As stated by Holmqvist et al. (2023) [25], participants with older ages are likely to have droopy eyelids, cataracts, macular degeneration or many other eye ailments. This ultimately results in eye-tracking recordings with either data losses (gaps), altered eye gaze/movements or a compromise in data quality.
- **Eye-blinks:** Considered as the most common data loss and systematic error factor, which means that blinks create big gaps as well as outliers in all trials. This will later on be fixed with processing.
- **Device location:** Changes in the placement of any of both devices affect data quality and, in some extreme cases, it can also happen to make a recording worthless, specially for eye-tracking.

Knowing the main factors for missing data chunks and poorer data quality, if a trial is set to be used in the analysis, it has to meet the following criteria:

1. **Data loss:** Preprocessed data for a given trial cannot have a data loss greater than 40%, meaning that the data must contain less than 40% NaN values.
2. **Delays:** Maximum permitted delay within the number of eye-tracking data samples and their respective timestamps is 0.5 seconds, while the maximum allowable delay between speech and eye-tracking data is 1 second.
3. **File association:** Trials consist of two distinct files, one for speech data and another one for eye-tracking data. If any of the files is missing, the trial isn't valid.

4. **Non-empty files:** All files must contain non-empty variables, which sometimes is the case for speech data.

2.5.3 Processing

Now that both the raw gaze and pupil size data have been preprocessed (in order to diminish the negative effect of blinks, non-uniform sampling, artifacts and data losses), next step consists of defining how to transform the valid raw samples (those that have not been discarded after preprocessing) into separate pupil diameter and fixation duration time signals.

Pupil diameter

As mentioned earlier in Section 2.5.1, the preprocessed pupil size data solely consists of ”valid” or ”non-erroneous” samples. Consequently, it is expected that the number of preprocessed samples will be lower than that of raw samples. To proceed with the processing of pupil size data from a trial, it is necessary that the proportion of preprocessed samples containing missing values (NaNs) is below 40% of the total number of samples, as previously explained on section 2.5.2. Taking advantage of the fact that each trial includes pupil size data from both eyes, the pupil diameter signal selected for a trial will be derived from the eye with the fewer number of NaNs after the preprocessing step.

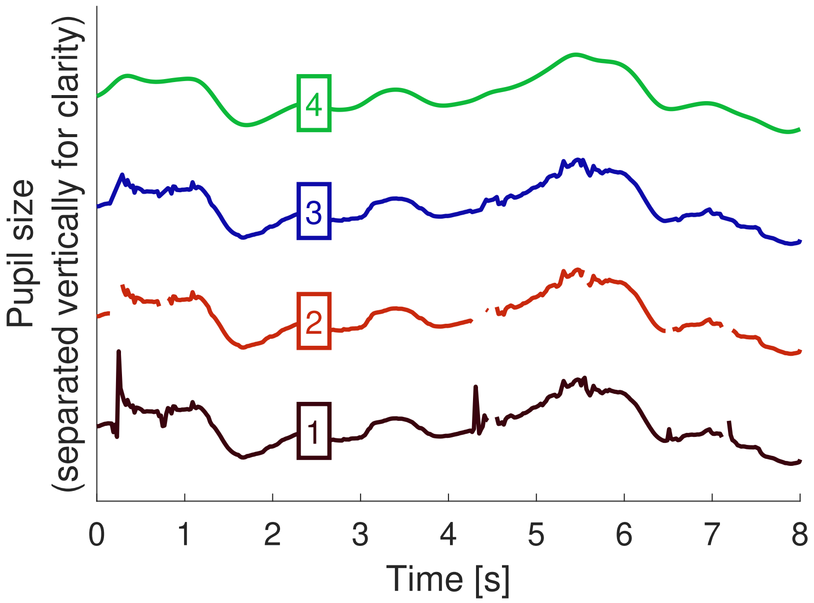


Figure 2.5: Example of pupil size data processing steps. Raw data, marked as **1**, is comprised of gaps, blinks and edge artifacts. Preprocessed data, marked as **2**, has excluded blink and edge artifact samples. Gaps are then interpolated, marked as **3**. And lastly, the data is low-pass filtered, marked as **4**.

In order to transform the preprocessed signal into a continuous pupil diameter signal, the first step is to linearly interpolate the chunks of missing data so that the signal doesn’t contain NaN values, similarly to what Winn et al. (2018) [21] described also as the next step after de-blinking.

Secondly, the data is low-pass filtered at 2 Hz (filter duration of 0.5 s) mainly due to the fact that Klingner et al. (2011) [26] found out that any pupil fluctuations faster than 10 Hz are not correlated accross eyes, meaning that any pupil response faster than 10 Hz should not be analyzed further as it probably is the result of artifacts, blinks or any other source of erroneous data. Low-pass filtering creates artifacts at the start and the end of the filtered signal (big peak and dip respectively) because the low-pass window values go from 0 to 1, this is taken care for by assigning the value of those peak/dip artifact samples as the Mean Pupil Dilation (MPD) of each trial (which won't affect as much to the overall response).

A visual representation of the overall processing from the raw data to the pupil size signal can be seen in fig. 2.5. The final processed signal has a distinct smoothness as well as an absence of any data losses, which indicates that all processing techniques have successfully turned the raw signal into a much more comprehensive and recognizable signal when compared to the results from any other pupillometry studies.

Baseline correction

Pupil dilation is commonly measured by assessing the change in pupil size relative to the time just before a stimulus is presented, rather than reporting the absolute pupil size [27]. Usually, baseline pupil size typically varies among participants, fluctuates over time within individuals, and gradually decreases during a testing session. These sources of variability at a trial-level as well as at a participant-level, can easily be solved by making sure that the number of participants and the duration of the trials were large enough.

In this thesis, the baseline is calculated separately for each speaking or listening window. This approach ensures that each response is equally estimated based on its own baseline, allowing us to gather these individual responses and subsequently compare them across different conditions. The chosen duration of baseline was of 0.5 seconds, as it is long enough to mitigate the possible effect of blinks and also short enough not to interfere with the previous window's completion. Knowing that D is the absolute pupil size (also known as diameter), the baselined pupil size, $D_{baselined}$, is then calculated as,

$$D_{baselined}(t) = D(t) - B(t), \tag{2.6}$$

being B the baseline calculated as an average,

$$B(t) = \frac{1}{n}\sum_{i=1}^{n} b_i = \frac{b_1 + b_2 + \cdots + b_n}{n}, \tag{2.7}$$

with n being the number of samples contained inside the baseline time interval, ranging from 0.5 seconds prior to the window's start until the start of the window itself.

Fixation duration

Eye movements and cognitive load have been suggested to be correlated or linked in a recent novel study by Cui et al. (2023) [28]. The study tends to suggest that the neural activity in brain regions that handle eye movements are modulated to some extent with listening effort. In order to extract fixation duration from the preprocessed gaze data, the methodology from the aforementioned paper by Cui et al. (2023) [28] was replicated. The data that will be utilized to estimate fixation duration, corresponds to the regions of preprocessed 3D gaze samples previously classified as fixations in section 2.5.1, by using the ”PUPILS” pipeline [24].

Fixation duration is calculated, sample by sample, as the number of previous and posterior samples whose euclidean distance with the current sample is less than or equal to the radius of a sphere (in the end it is divided by the sampling frequency, so that its unit is seconds). This theoretical sphere represents the fovea or foveal area, the central region of the retina in the human eye that is key for visual perception and also fixation. The fovea is characterized by a high density of cone cells, which are specialized photo-receptor cells that are responsible for color and detailed visual acuity. Due to its small size and concentrated cone cell distribution, the fovea enables the capture of fine details and provides a high level of visual clarity within a limited field of view.

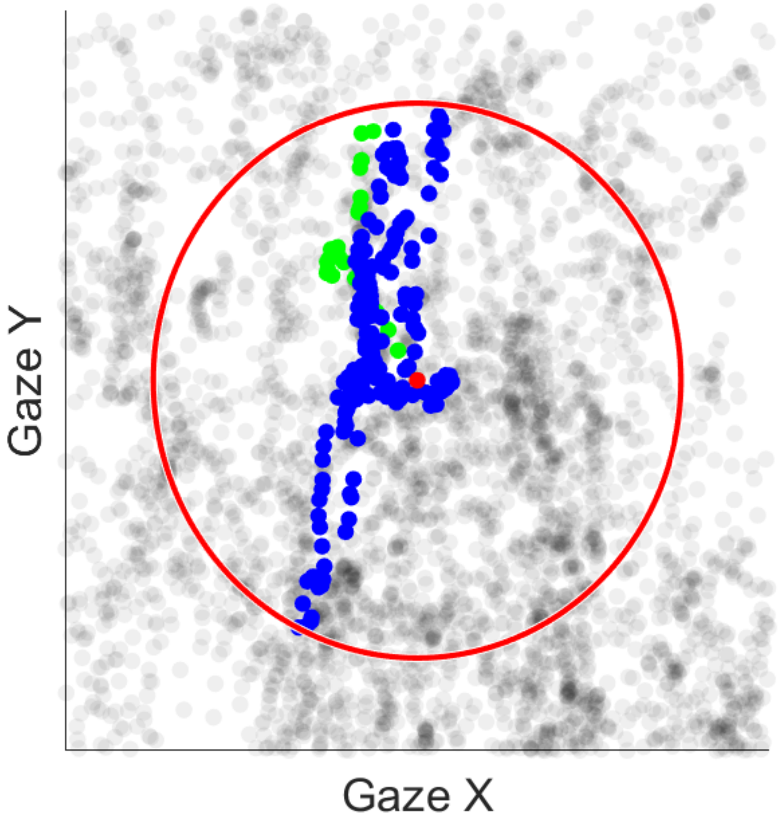


Figure 2.6: Fixation duration calculation for one gaze sample (current sample). The sum of all contiguous previous and posterior samples with a distance less than the radius (given by the foveal area of 1.5°), divided by the sampling frequency, is the fixation duration for the current sample (in this case is 4.16 seconds or 208 contiguous samples). Other gaze samples are shown in gray.

The width in degrees of the fovea is needed to calculate the radius of the aforementioned theoretical sphere. The measure of the foveal area's width was first introduced by Helmholtz in 1868, by defining the concept of the thumb's rule, a rule that suggests that the visual angle formed by the thumb held at arm's length corresponds to a small integer value. An initial study by Robert P. O'Shea in 1991 [29] concluded that the visual angle for the width of the thumb is around 2°, with the thumbnail width estimated at approximately 1.5°. For this reason, the value used in the processing of this thesis as the fovea visual angle is 1.5°. In fig. 2.6, an example of the calculation of the fixation duration from one sample in particular is shown.

Once the fixation duration has been calculated, the signal is low pass filtered with a hamming window at 2 Hz (duration of 0.5 s). This is done with the help of a zero-phase digital filter that ignores NaNs. An example of the fixation duration processing can be found in fig. 2.7, in which it can be easily seen that blink and saccadic regions will not be used on the estimation of the fixation duration.

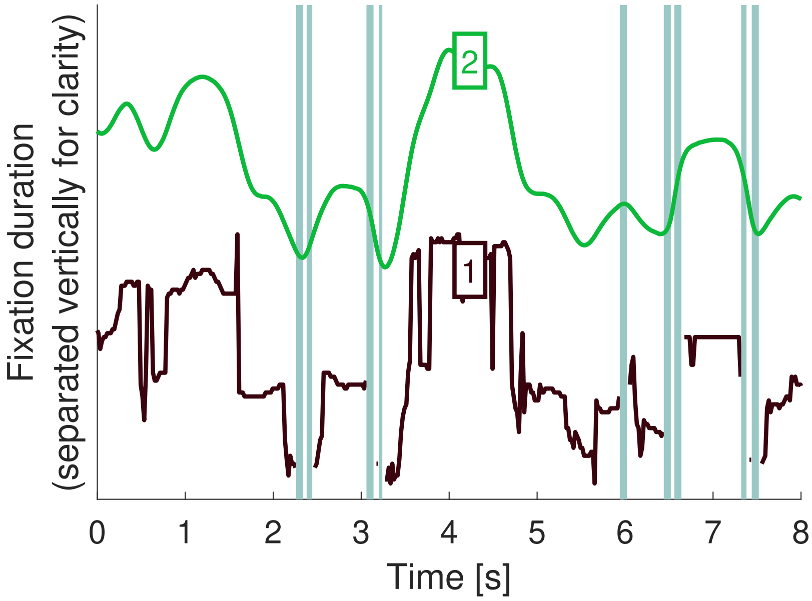


Figure 2.7: Example of fixation duration data processing steps. Gaze data is preprocessed first (as explained in section 2.5.1), the regions that are not marked as fixations, which are saccades and blinks (shown as turquoise vertical rectangles in the figure), will be ignored during the raw fixation duration, marked as **1** (which explains the gaps in the raw fixation duration). Then, the fixation duration is low-pass filtered, marked as **2**.

2.5.4 Speech data

The desired outcome from the microphone recordings are both the listening and speaking windows for each subject on each trial. In order to do so, the audio tracks from each subject (coming through the head-worn microphones mentioned in section 2.1.1) were analyzed individually for each trial using the Communicative State Classification (CSC) algorithm described in a study by Sørensen et al. (2021) [30]. Conversations are fundamentally structured in turns that alternate (or not) between speakers, this transition from one talker to another is named floor-transfer. The timing involved in turn-taking has been termed as the floor-transfer offset (FTO), which is defined as the interval from when a person stops talking to when another person starts talking.

The algorithm [30] was responsible of labelling short 4 ms segments as either speech (1) or no speech (0), with the use of Voice Activity Detection (VAD), based on the Root Mean Square (RMS) value of each segment. This 250 Hz binary activity array output was then used to classify the conversation, thanks to the CSC algorithm, into six possible time groups or windows:

1. **Utterances:** Also named interpausal units (IPUs), these are windows that contain connected speech with silences no bigger than 180 ms.
2. **Turns:** Duration of IPUs by one talker, delimited by floor-tranfers.
3. **Gaps:** Silence during floor-transfers.
4. **Pauses:** Silence in turns (no floor-transfer).

5. **Overlap-Between:** Overlapping speech during a floor-transfer.
6. **Overlap-Within:** Fully overlapped speech occurring on someone's turn.

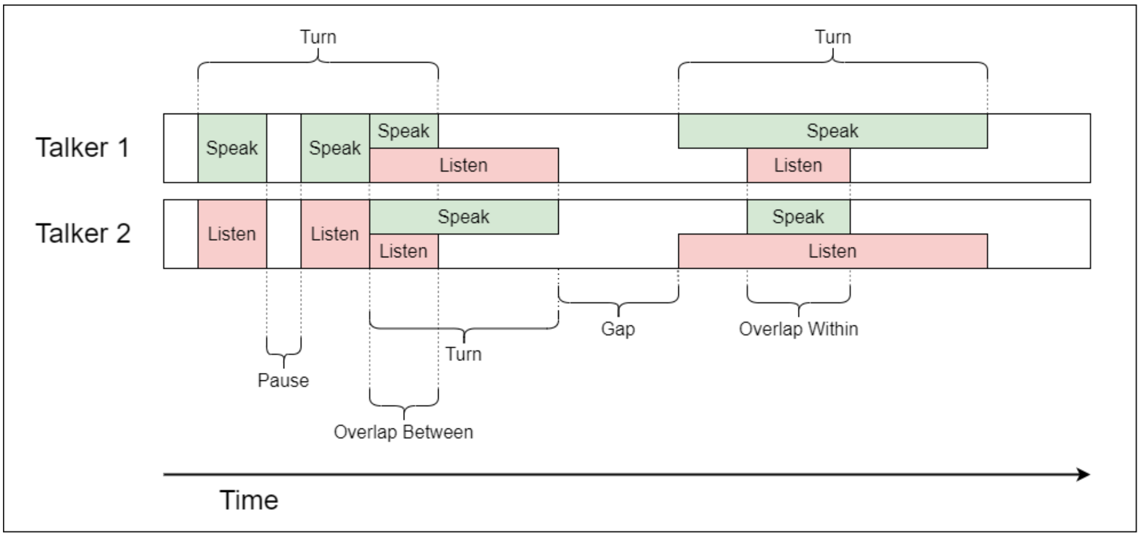


Figure 2.8: Simple sketch of the output of the CSC algorithm by Sørensen et al. (2021) [30], which classifies gaps, pauses, turns, overlaps (within and between) as well as utterances or IPUs (speaking or listening).

The utterances or IPUs were then used to define either speaking or listening windows, depending on which talker belongs to which channel for a specific trial. In fig. 2.8, there is a visual representation of all the possible groups in a conversation, with the addition of the classification of utterances or IPUs as either speaking or listening. Note that, all speaking windows translate to listening windows for the other talker and viceversa.

The CSC algorithm [30] was applied to each non-rejected trial for both AMEND I and AMEND II, with the duration of the windows shown as histogram plots in both section 3.1.1 and section 3.1.2 respectively. Additionally, their averages are also displayed in the "**Raw**" row in both table 3.1 and table 3.2. After a preliminary analysis, it becomes apparent that the duration for speaking and listening (IPUs) windows, is relatively shorter in our analysis when compared to an example study by Levinson et al. (2015) [31] that shared the maximum silence duration within IPUs of 180 ms.

Speaking and listening windows must be large enough in order to be able to fully capture both the response of the pupil (dilation) as well as the response of the eye gaze or movements (fixations). In order to do so, the initial raw speaking and listening windows from the outcome of the algorithm by Sørensen et al. (2021) [30], were further processed in the following sequence (shown in fig. 2.9):

- Merge any consecutive windows from the same type (speaking or listening) if there's a time interval between them smaller or equal to 0.3 seconds.
- Discard any windows with a duration smaller or equal to 0.5 seconds.
- Merge windows again, this time allowing for a time interval smaller or equal to 2 seconds.
- Discard again, only for those windows with a duration smaller or equal to 1 second.

- Split overlaps-within into new separate windows and ignore overlaps-between, keeping only the window with the latest active talker.

The reasoning for selecting these different thresholds in the processing of speaking and listening windows is that, given the fact that the aforementioned algorithm by Sørensen et al. (2021) [30] works with audio as an input, it is expected that not all of the outcome windows directly correspond to speech. Many undesired participant-related sources of noise happened during the trials, some examples are laughter, coughing, expressions for agreement such as ”ja”, ”uh-huh” and many more. Literature by Levinson (2015) [31], Schegloff (2000) [32] and Gravano (2012) [33] have managed to define different categories in order to describe sources of ”noise” in the overlap-within windows, such as, simultaneous start from both talkers, turn-taking attempts, uncompleted turns and some more. Processing both speaking and listening windows ensures that participant responses are fully captured while also effectively filtering out undesired rapid windows.

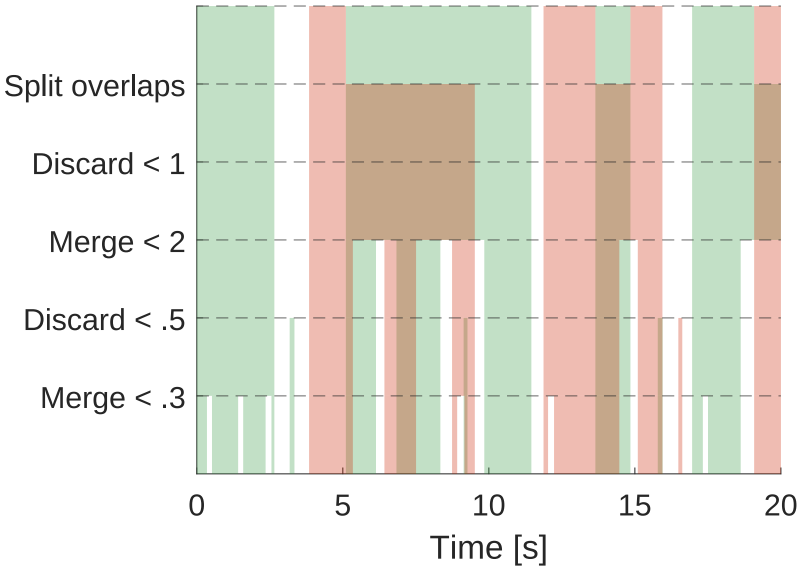


Figure 2.9: Example of the post-processing sequence, from bottom to top, applied to both speaking and listening windows obtained using the CSC algorithm by Sørensen et al. (2021) [30]. The initial **raw** windows, on the bottom, undergo the aforementioned processing sequence to finish as the **processed** speaking and listening windows on the top. Color legend is shown both in fig. 3.1 and in fig. 3.2

2.5.5 Pupillometry features

In order to analyze the pupil size results after processing, it’s good practice to define some key features that can represent many different aspects of the pupil response over time, providing an easy measure for each window to relate and compare between different TPs and conditions. The features chosen for further analysis were:

- **Mean Pupil Dilation (MPD):** Estimated as the average pupil size value within a specific window. It provides a reliable representation of the general level of arousal or cognitive load experienced by the individual. As indicated in an article by Ahmad et al. (2020) [34], not only MPD was found to be the most prominent eye feature but also it concluded that higher mean pupil dilation values generally indicate increased arousal or cognitive load, while lower values suggest reduced arousal or cognitive demand.

- **Slope:** First coefficient of a linear regression fitted using the pupil size values within a specific window. It captures the steepness or incline of the response, reflecting the speed or the rate of change (and its intensity).
- **Peak Pupil Dilation (PPD):** Maximum pupil size value within a specific window. This feature indicates moments of overall heightened attention, emotional arousal, or cognitive engagement in a window. It provides insights into the salience or significance of the pupil response from a TP within a certain window and noise condition or hearing aid setting.

By considering these three pupil size features in the analysis of the results, a comprehensive understanding of the different pupil dynamic response can be gained. Enabling a simple assessment within different windows of the cognitive load and cognitive processing for different conditions. These features will later on be compared against each other in search of statistically significant differences.

2.5.6 Linear Mixed Modelling

Linear Mixed Models (LMMs), also known as mixed-effects models or hierarchical linear models, are a powerful statistical framework for analyzing data that exhibit both fixed and random effects, as explained in an article by Singmann et al. (2019) [35]. LMMs expand from traditional linear regression models by incorporating random effects, which account for the variability between the groups or clusters in the data.

In this case, this framework will be used with the measures obtained from the pupil features estimated from samples belonging to different windows, as defined in section 2.5.5, all data from the features corresponding to AMEND I and AMEND II will be collected in two different tables. The variables (table columns) of the table are: Subject (TP associated number), activity (window being either speaking or listening), noise condition (or hearing aid setting in AMEND II) and lastly, the three estimated features within the window: MPD, Slope and PPD.

To fit the LMM to the data from the features, MATLAB's [15] built-in function `fitlme()` was used with the default optional parameters. The input arguments where the aforementioned tables and a formula which defined the fixed effects and the random effects for the fitting of the model. The different formulas can be seen in listing 2.1, with the feature of interest first (as the response variable), the activity and noise condition or hearing aid setting as the fixed effects for the intercept and, lastly, the subject as each level of the grouping of the random effect for the intercept.

The intercept is the reference from which the model will estimate the differences with, in our case, we chose our intercept to be the speaking or listening activity combined with either the quiet noise condition or the un-aided hearing aid setting.

```
% Formulas to differentiate between noise conditions
form_N_MPD='MPD~Activity*Noise+(1|Subject)';
form_N_Slope='Slope~Activity*Noise+(1|Subject)';
form_N_PPD='PPD~Activity*Noise+(1|Subject)';

% Formulas to differentiate between hearind aid settings
form_S_MPD='MPD~Activity*Setting+(1|Subject)';
form_S_Slope='Slope~Activity*Setting+(1|Subject)';
form_S_PPD='PPD~Activity*Setting+(1|Subject)';
```

Listing 2.1: Formulas for fitting LMMs in MATLAB [15].

3 Results

This chapter will present the results from both AMEND I and AMEND II, these results were evaluated and analyzed based on the hypothesis presented in section 1.3. The main objective of this chapter is to present all processed data from the experiments in order to give out a simple and clear representation of the findings. These will later on be used to form a discussion to describe the behavioral responses and to what extent are they dependant to any of the different conditions or groups of participants.

3.1 Communication dynamics results

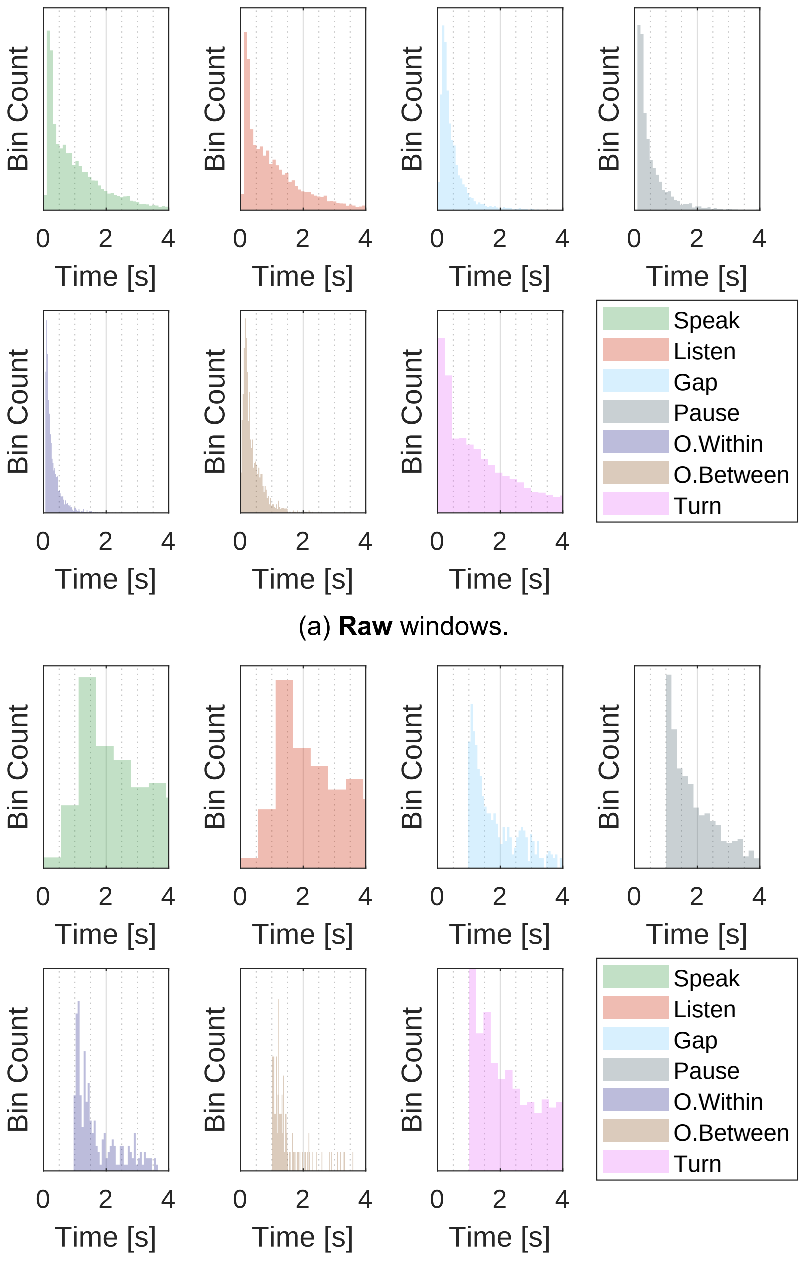
The **raw** windows obtained with the method described previously in section 2.5.4, belong to different types. The utterances are classified into either speaking or listening, depending on the actions of each speaker. Additionally, each trial will then contain (or not) different windows named as: Speak, Listen, Gap, Pause, Overlap-Within, Overlap-Between or Turn.

The windows' durations shown in both section 3.1.1 (AMEND I) and section 3.1.2 (AMEND II), reveal a great difference on their time distribution when comparing both the **raw** to the **processed** windows, seen in fig. 3.1 and fig. 3.2 respectively. This duration difference, corresponds to 4 seconds approximately, indicating that once speaking and listening windows are processed (with the method defined in section 2.5.4), they will then be large enough in size to fully capture pupil and fixation responses to each event (speaking or listening) from each participant.

3.1.1 AMEND I

Table 3.1: Average duration (in seconds) from the histograms shown in fig. 3.1.

Avg. duration [s]	Speak	Listen	Gap	Pause	OLW	OLB	Turn
Raw	1.09	1.12	0.49	0.55	0.30	0.38	1.99
Processed	5.12	5.33	2.00	2.36	1.98	1.50	7.06



(a) **Raw** windows.

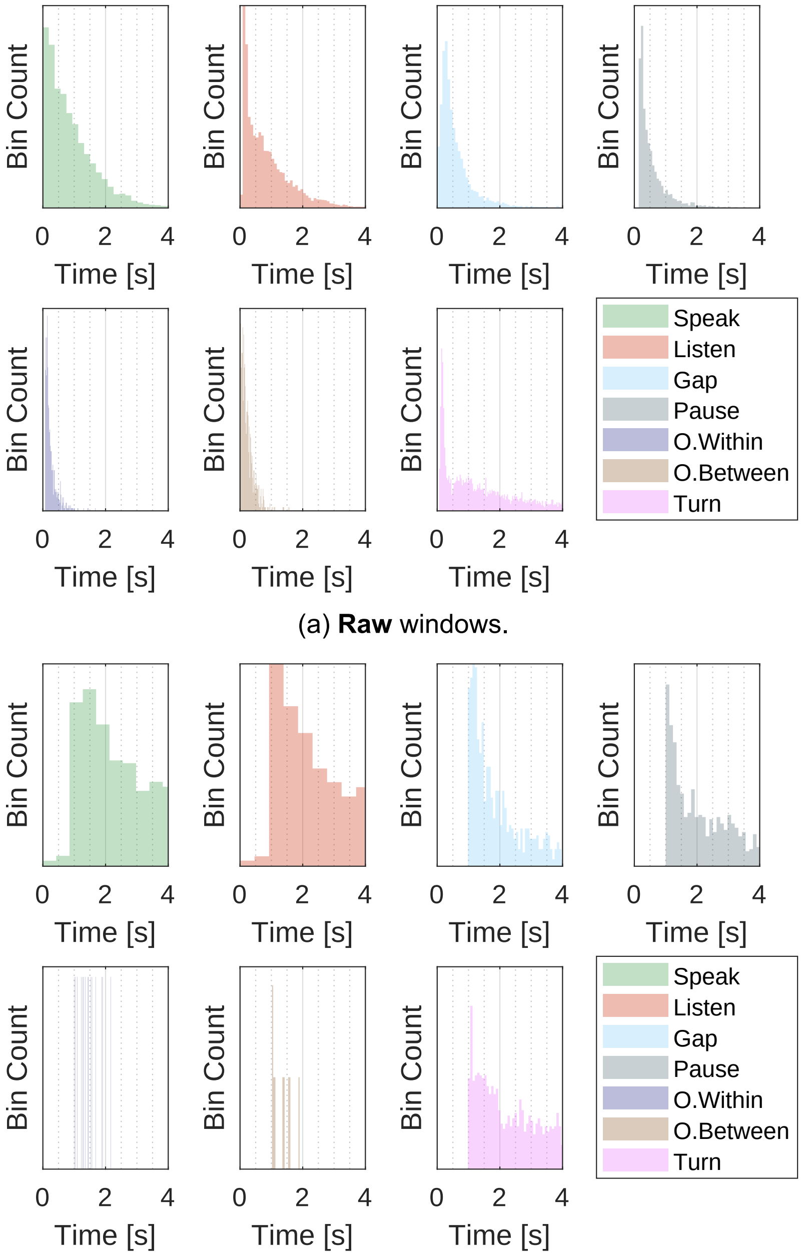
(b) **Processed** windows.

Figure 3.1: Histograms of the duration from both the **raw** and **processed** CSC outcome measures (different windows) in AMEND I, obtained by the method described in section 2.5.4. The average duration of each type of window is shown on table 3.1.

3.1.2 AMEND II

Table 3.2: Average duration (in seconds) from the histograms shown in fig. 3.2.

Avg. duration [s]	Speak	Listen	Gap	Pause	OLW	OLB	Turn
Raw	0.90	0.88	0.60	0.59	0.25	0.23	2.12
Processed	4.92	4.657	1.99	2.66	1.46	1.67	6.654



(a) **Raw** windows.

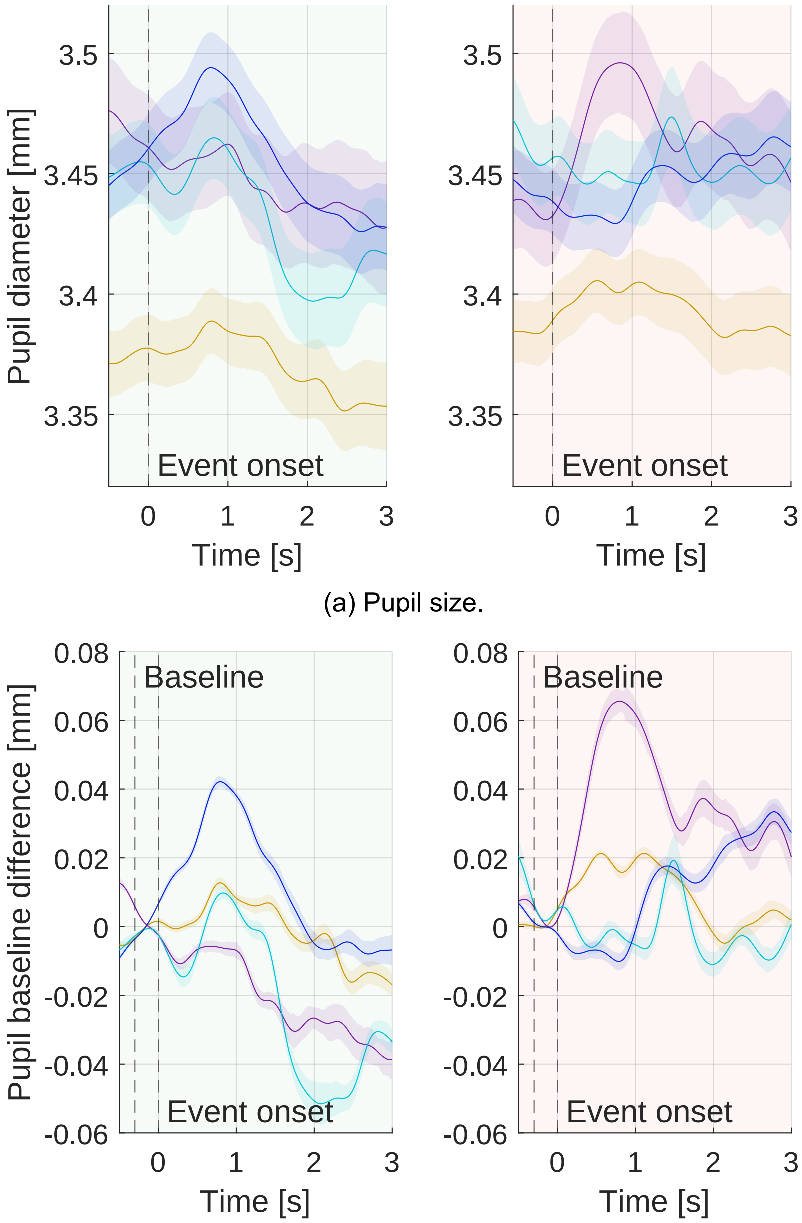
(b) **Processed** windows.

Figure 3.2: Histograms of the duration from both the **raw** and **processed** CSC outcome measures (different windows) in AMEND II, obtained by the method described in section 2.5.4. The average duration of each type of window is shown on table 3.2.

3.2 Pupil size results

This section presents the pupil size results derived from all the non-rejected trials in AMEND I and AMEND II. All these processing steps are described in details in chapter 2. Results are presented in the form of pupil diameter over time (AMEND I in fig. 3.3 and AMEND II in fig. 3.6 and fig. 3.8) and also in the form of grouped pupil size features (AMEND I in fig. 3.5 and AMEND II in fig. 3.9 and fig. 3.10). The results are separated by different test conditions and/or by hearing aid settings (only for HI in AMEND II).

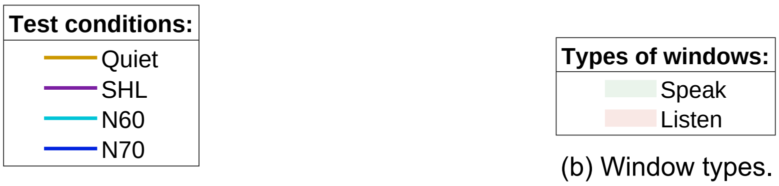
3.2.1 AMEND I



(a) Pupil size.

(b) Baselined pupil size.

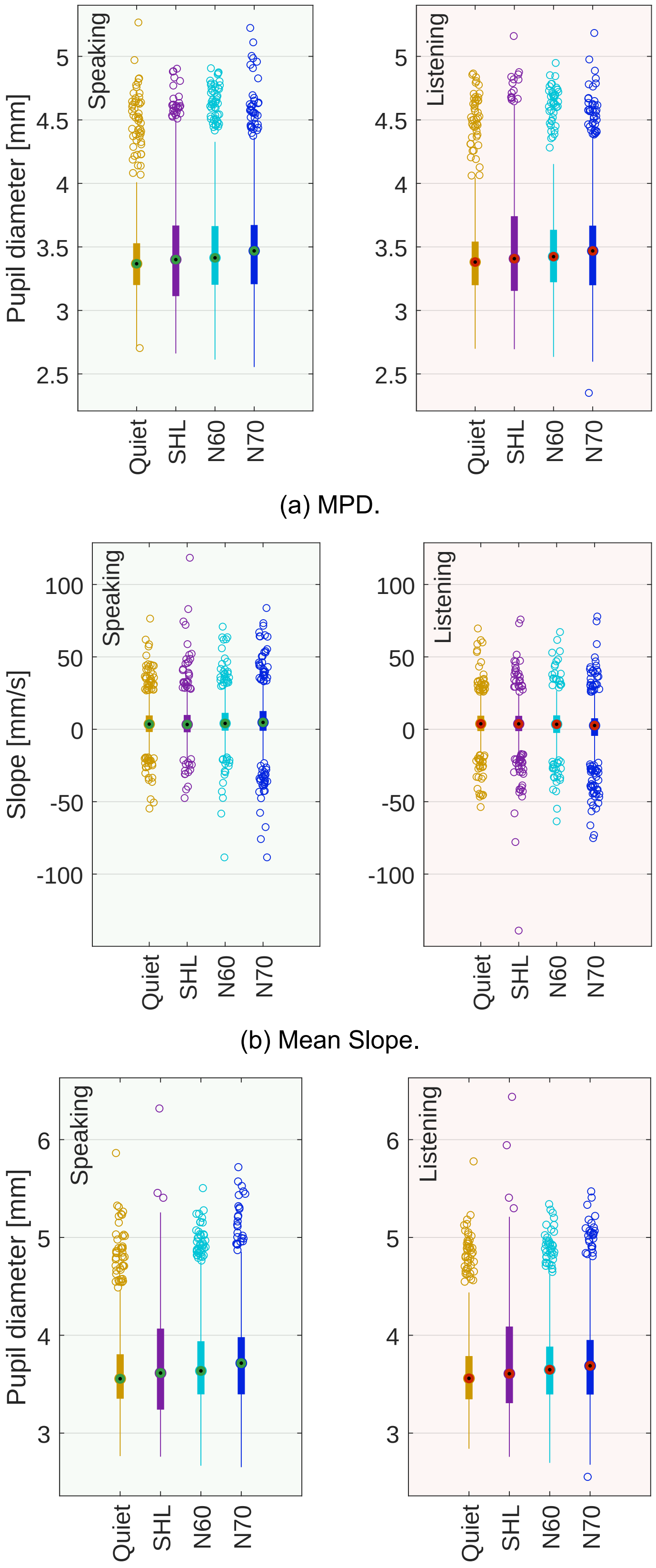
Figure 3.3: Global pupil size results averaged across participants and separated different test conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND I. Color legend is shown on fig. 3.4.



(a) Test conditions.

(b) Window types.

Figure 3.4: Color legend in AMEND I.



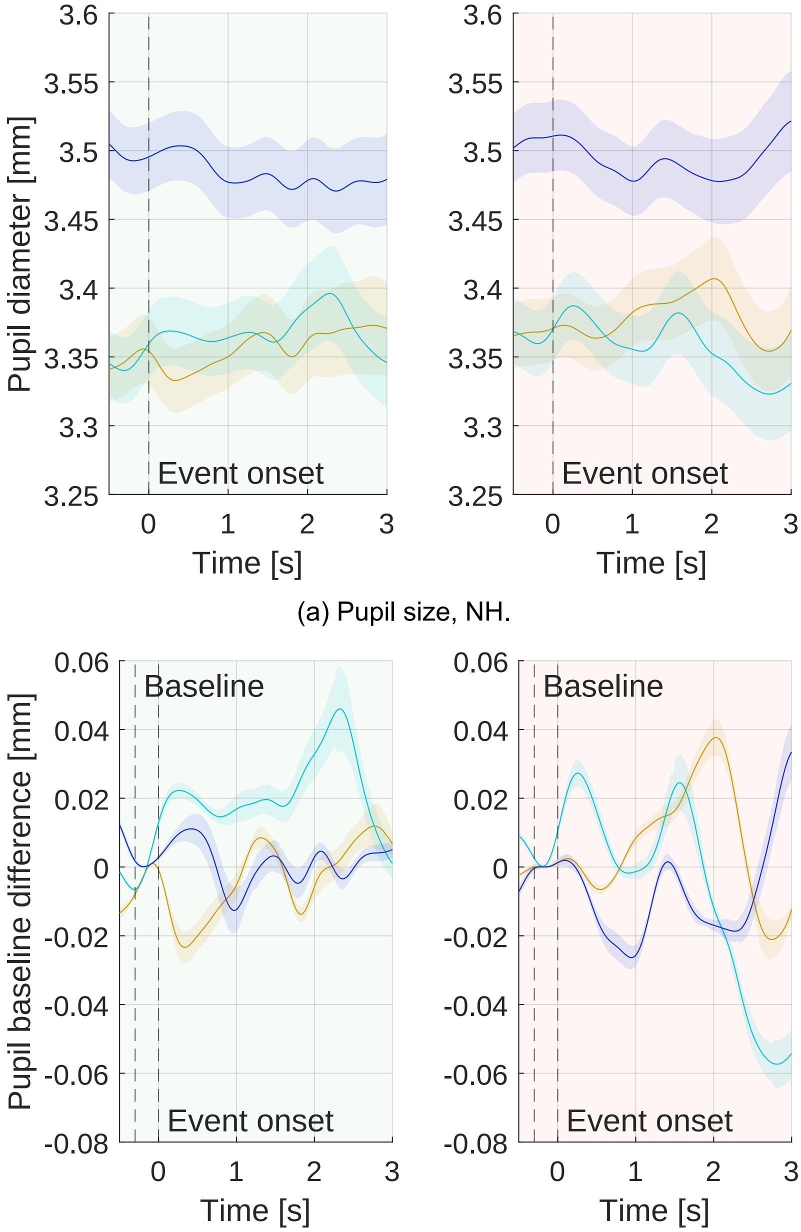
(a) MPD.

(b) Mean Slope.

(c) PPD.

Figure 3.5: Pupil size features (MPD, Slope and PPD) in AMEND I separated by different test conditions, calculated from the samples within the same time interval as in fig. 3.3b (from 0 to 3 s after event onset). Features belong to both speaking and listening windows (left and right panels, respectively). Color legend is shown on fig. 3.4.

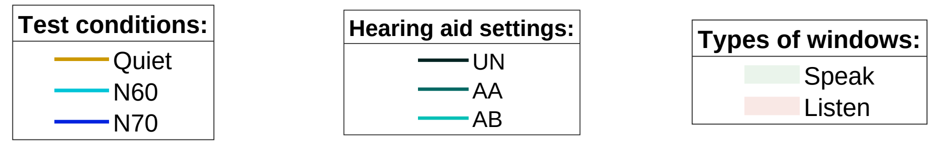
3.2.2 AMEND II



(a) Pupil size, NH.

(b) Baselined pupil size, NH.

Figure 3.6: Normal hearing (NH) pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

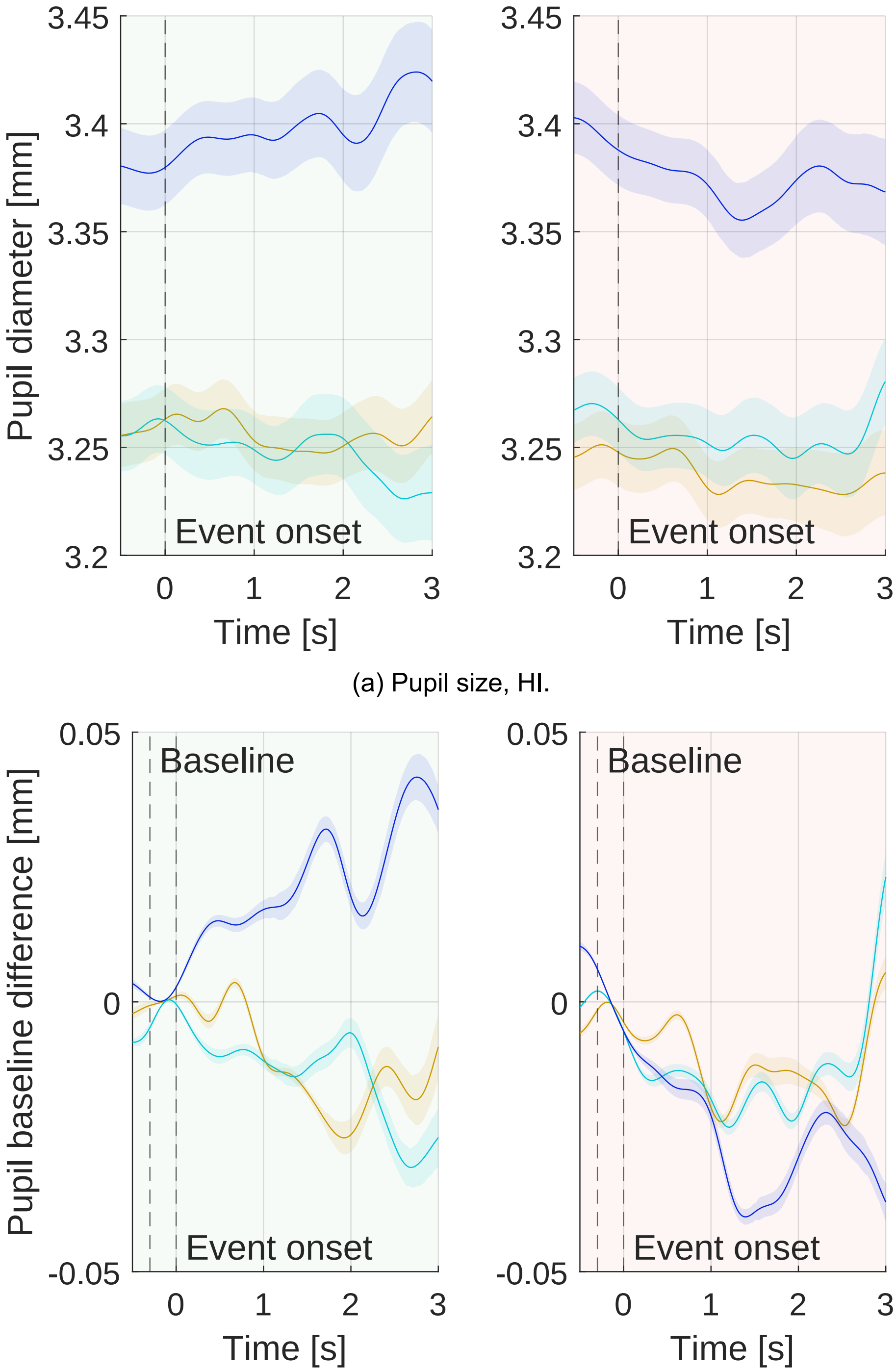


(a) Test conditions.

(b) Hearing aids settings.

(c) Window types.

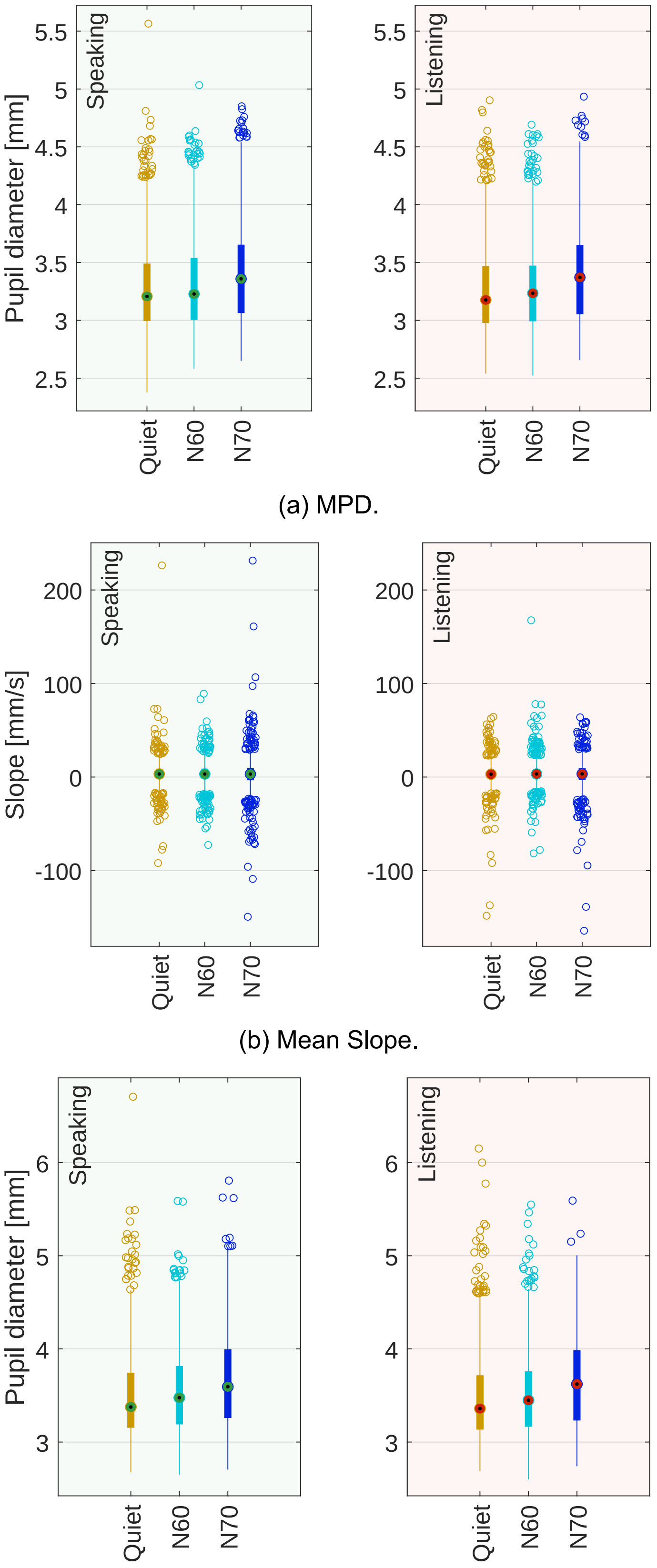
Figure 3.7: Color legend in AMEND II.



(a) Pupil size, HI.

(b) Baselined pupil size, HI.

Figure 3.8: Hearing impaired (HI) pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

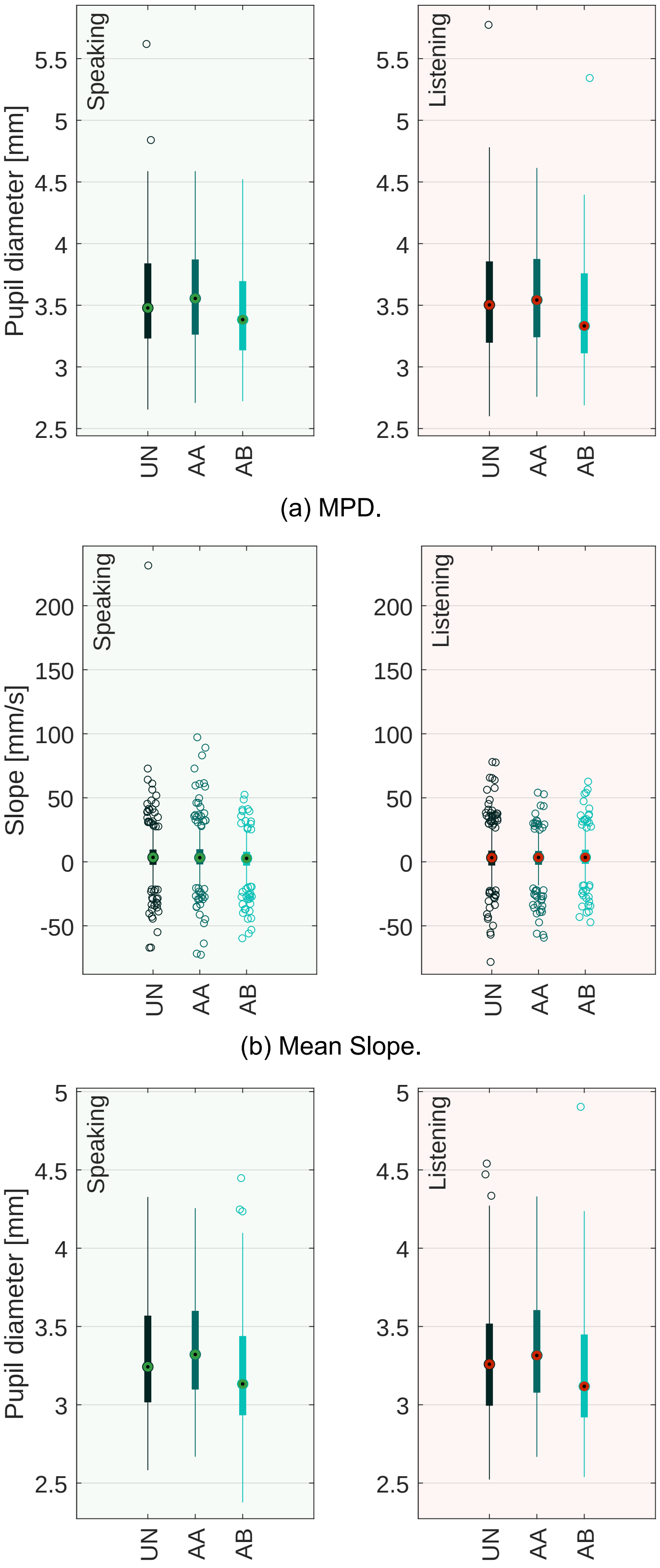


(a) MPD.

(b) Mean Slope.

(c) PPD.

Figure 3.9: Pupil size features (MPD, Slope and PPD) in AMEND II separated by noise conditions, calculated from the samples within the same time interval as in fig. 3.6b (from 0 to 3 s after event onset). Features belong to both speaking and listening windows (left and right panels, respectively). Color legend is shown on fig. 3.7.



(a) MPD.

(b) Mean Slope.

(c) PPD.

Figure 3.10: Pupil size features (MPD, Slope and PPD) for HI in AMEND II separated by hearing aids settings (as described in section 2.4), calculated from the samples within the same time interval as in fig. 3.8b (from 0 to 3 s after event onset). Features belong to both speaking and listening windows (left and right panels, respectively). Color legend is shown on fig. 3.7.

3.2.3 Statistic analysis of pupil size features

To improve the characterization of the pupil features, it is required to conduct an statistical analysis. A mere representation of the features, such as in the form of box-and-whisker plots (as exemplified in fig. 3.5), is insufficient to fully quantify and characterize the features. The chosen p-value threshold, typically set at **0.05** (or 5e-2), is underscored in the literature [35] as an essential step when performing a LMM analysis.

Evaluating the p-value against this threshold holds key importance in determining the statistical significance and meaningful differences between a given condition and the intercept while minimizing the risk of Type I errors. Another crucial aspect in the LMM analysis is the presence of zero (or not) within the confidence bound estimates. This determination is based on evaluating the signs of the lower and upper bounds. The absence of zero within the confidence bound indicates statistical significance, providing evidence of a genuine effect. On the other hand, the presence of zero suggests potential non-significance of the effect.

Therefore, LMMs are used exactly as described in section 2.5.6, in search of significant differences between intercepts and conditions, resulting in the following metrics:

AMEND I:

In AMEND I feature analysis, the intercept or reference condition corresponds to the quiet condition in either speaking or listening windows. The output metrics of the LMM for various features, namely MPD, Slope, and PPD, are presented in table 3.3, table 3.4, and table 3.5 correspondingly.

Table 3.3: LMM metrics from the MPD feature (as represented in fig. 3.5a) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND I.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	3.3835	0.0755	44.801	2389	1.03e-318	3.2354	3.5316
	SHL	0.0717	0.0116	6.1679	2389	**8.10e-10**	0.0489	0.0945
	N60	0.0514	0.0111	4.6366	2389	**3.73e-06**	0.0297	0.0732
	N70	0.0858	0.0101	8.5604	2389	**1.98e-17**	0.0661	0.1054
Listening	Quiet	3.4023	0.0763	44.576	2417	3.1e-317	3.2526	3.552
	SHL	0.0786	0.0116	6.7421	2417	**1.94e-11**	0.0557	0.1015
	N60	0.0415	0.0110	3.7649	2417	**1.70e-04**	0.0199	0.0632
	N70	0.0591	0.0101	5.8573	2417	**5.34e-09**	0.0392	0.0788

Table 3.4: LMM metrics from the Slope feature (as represented in fig. 3.5b) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND I.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	4.3982	0.8593	5.1185	2389	3.32e-07	2.7132	6.0831
	SHL	0.0551	0.9447	0.0583	2389	0.9535	-1.7974	1.9076
	N60	1.1279	0.8937	1.262	2389	0.2071	-0.6247	2.8805
	N70	1.2947	0.8205	1.5784	2389	0.1146	-0.3139	2.904
Listening	Quiet	4.1535	0.5698	7.2898	2417	4.18e-13	3.0362	5.2708
	SHL	-0.6554	0.8928	-0.7341	2417	0.4629	-2.4062	1.0953
	N60	-0.7495	0.8377	-0.8947	2417	0.3708	-2.3922	0.8932
	N70	-3.1889	0.7914	-4.0296	2417	**5.76e-05**	-4.7407	-1.6371

Table 3.5: LMM metrics from the PPD feature (as represented in fig. 3.5c) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND I.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	3.6055	0.0878	41.082	2389	1.46e-279	3.4334	3.7776
	SHL	0.0901	0.0161	5.5908	2389	**2.52e-08**	0.0584	0.1216
	N60	0.0511	0.0154	3.3213	2389	**9.09e-04**	0.0209	0.0812
	N70	0.1260	0.0139	9.081	2389	**2.19e-19**	0.0988	0.1532
Listening	Quiet	3.62	0.0879	41.166	2417	3.57e-281	3.4475	3.7924
	SHL	0.1017	0.0157	6.4744	2417	**1.15e-10**	0.0709	0.1325
	N60	0.0415	0.0149	2.79	2417	**5.31e-03**	0.0123	0.0706
	N70	0.0892	0.0136	6.573	2417	**6.02e-11**	0.0626	0.1158

From these tables, the p-values that are smaller than the aforementioned threshold are marked in **bold**. Meaning that, for the MPD and PPD features, each one of the noise conditions (SHL, N60 and N70) are significantly different from the reference (Quiet) both for speaking and listening. As for the slope feature, the only significant difference can be found in listening windows between the ”N70” noise condition and the ”Quiet” reference.

AMEND II:

For the noise condition analysis in AMEND II, the intercept or reference condition also corresponds to quiet test condition in either speaking or listening windows (same as in AMEND I). The LMM output metrics for the MPD, Slope and PPD features are shown in table 3.6, table 3.7 and table 3.8. Once again, p-values smaller than the threshold (0.05) are marked in **bold**.

Table 3.6: LMM metrics from the MPD feature (as represented in fig. 3.9a) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	3.3013	0.0931	35.445	2289	9.44e-220	3.1186	3.4839
	N60	0.0214	0.0112	1.9078	2289	0.0565	-0.0006	0.0434
	N70	0.1259	0.0112	11.208	2289	**2.01e-28**	0.1039	0.1479
Listening	Quiet	3.3025	0.0907	36.405	2224	4.94e-228	3.1246	3.4804
	N60	0.0075	0.0113	0.6579	2224	0.5106	-0.0148	0.0297
	N70	0.1313	0.0114	11.51	2224	**8.03e-30**	0.1089	0.1536

Table 3.7: LMM metrics from the Slope feature (as represented in fig. 3.9b) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	3.6234	0.6384	5.6759	2288	1.55e-08	2.3715	4.8752
	N60	-0.5301	0.9123	-0.581	2288	0.5613	-2.3191	1.259
	N70	-0.4697	0.9151	-0.5132	2288	0.6079	-2.2642	1.3249
Listening	Quiet	2.7673	0.5945	4.6548	2224	3.43e-06	1.6015	3.9332
	N60	1.5682	0.8481	1.849	2224	0.0646	-0.0950	3.2314
	N70	-0.2955	0.8529	-0.3465	2224	0.729	-1.968	1.377

Table 3.8: LMM metrics from the PPD feature (as represented in fig. 3.9c) separated by noise conditions and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Noise	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	Quiet	3.5514	0.116	30.629	2289	5.85e-173	3.3241	3.7788
	N60	0.0467	0.0145	3.2029	2289	**1.38e-3**	0.0181	0.0754
	N70	0.1681	0.0146	11.49	2289	**9.48e-30**	0.1394	0.1968
Listening	Quiet	3.5506	0.1124	31.595	2224	2.66e-181	3.3302	3.7709
	N60	0.0268	0.0149	1.7876	2224	0.07398	-0.0026	0.0562
	N70	0.1661	0.0151	11.022	2224	**1.52e-27**	0.1365	0.1956

For the MPD feature, significant differences were observed in both speaking and listening conditions in the ”N70” condition. However, no significant differences were detected in the slope feature. On the other hand, the PPD feature exhibited differences with the ”N60” and ”N70” conditions in the speaking windows, as well as a significant difference in the ”N70” condition during the listening windows.

As for the analysis of the hearing aid settings in AMEND II, the intercept or reference haring aid setting is the un-aided hearing aid setting (UN) both for speaking and listening windows. The LLMs metrics are shown in table 3.9, table 3.10 and table 3.11 for the MPD, slope and PPD features. Any p-value smaller than the threshold (0.05) is marked in **bold**.

Table 3.9: LMM metrics from the MPD feature (as represented in fig. 3.10a) separated by hearing aid settings and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Setting	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	UN	3.3146	0.0971	34.15	1159	1.94e-177	3.1241	3.505
	AA	-0.015	0.0171	-0.8784	1159	0.37992	-0.0486	0.0185
	AB	-0.0681	0.0171	-3.988	1159	**7.08e-05**	-0.1016	-0.0346
Listening	UN	3.3088	0.0976	33.89	1160	1.48e-175	3.1172	3.5003
	AA	0.0137	0.0174	0.7868	1160	0.4316	-0.0204	0.0478
	AB	-0.0528	0.0175	-3.0178	1160	**0.0026**	-0.0871	-0.0185

Table 3.10: LMM metrics from the Slope feature (as represented in fig. 3.10b) separated by hearing aid settings and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Setting	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	UN	3.7951	0.8193	4.632	1158	4.02e-06	2.1877	5.4025
	AA	0.2132	1.24	0.1719	1158	0.8636	-2.2198	2.6461
	AB	-2.0032	1.244	-1.6103	1158	0.1076	-4.4439	0.4375
Listening	UN	3.9449	0.7074	5.5765	1160	3.05e-08	2.5569	5.3328
	AA	-1.2864	1.0618	-1.2116	1160	0.2259	-3.3696	0.7968
	AB	0.0468	1.0685	0.0438	1160	0.965	-2.0495	2.1432

Table 3.11: LMM metrics from the PPD feature (as represented in fig. 3.10c) separated by hearing aid settings and belonging to either speaking or listening windows, extracted from data in AMEND II.

	Setting	Estimate	SE	t stat	DOF	p-value	Lower	Upper
Speaking	UN	3.5662	0.1005	35.475	1159	3.11e-187	3.3689	3.7634
	AA	-0.0261	0.0222	-1.1786	1159	0.2388	-0.0696	0.0174
	AB	-0.079	0.0221	-3.5721	1159	**3.68e-04**	-0.1224	-0.036
Listening	UN	3.5567	0.1024	34.72	1160	1.07e-181	3.3557	3.7577
	AA	-0.0025	0.0222	-0.1126	1160	0.9104	-0.0461	0.0411
	AB	-0.0572	0.0224	-2.5589	1160	**0.0106**	-0.1012	-0.0134

Both the MPD and PPD features indicated significant differences between the reference setting (UN) and the setting B (AB), both in speaking and listening windows. The slope, as seen previously with the analysis of the noise conditions, didn’t give out any significant difference.

In general, all features gave out significant differences, specially the MPD and PPD features which, compared to the slope, highlighted many more p-values lower than the threshold (0.05). These will later on be discussed on chapter 4.

3.3 Fixation duration results

This section presents the fixation duration results derived from the non-rejected trials in AMEND I and AMEND II. More specifically, it showcases the fixation duration in seconds over time, obtained within speaking or listening time windows. These measures are displayed in fig. 3.11 for AMEND I and also in fig. 3.13 and fig. 3.14 for AMEND II. The calculation and processing of these fixation duration signals over time is stated in section 2.5.3.

3.3.1 AMEND I

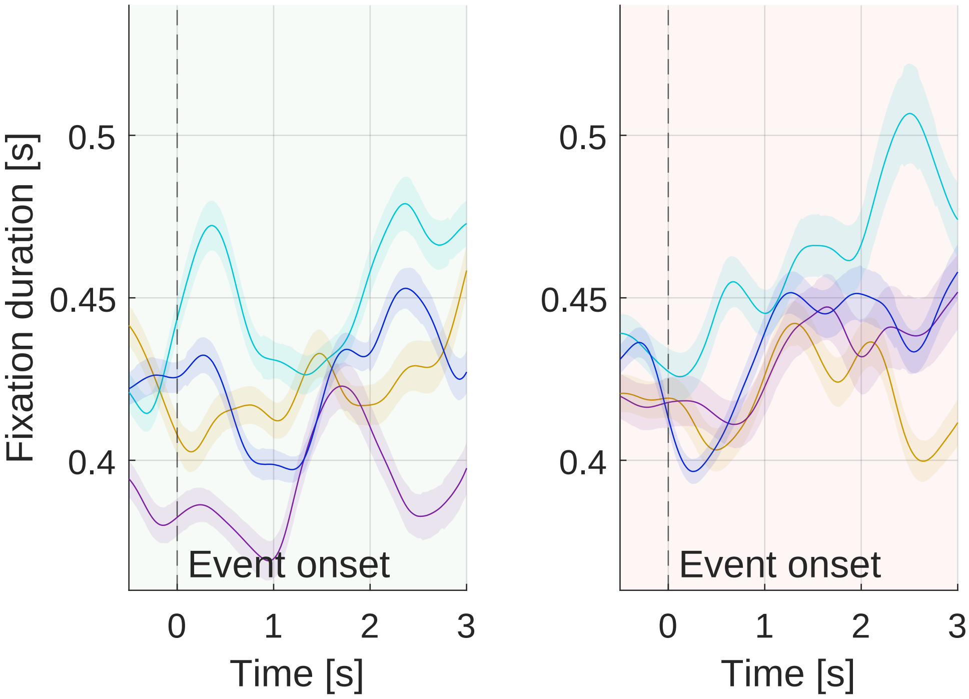
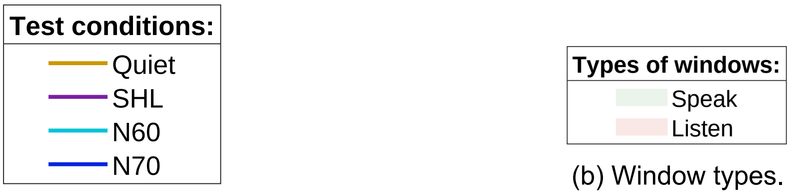


Figure 3.11: Fixation duration global results averaged across participants separated by different test conditions, with events being either speaking or listening windows in AMEND I. Color legend is shown on fig. 3.12.



(a) Test conditions.

(b) Window types.

Figure 3.12: Color legend in AMEND I. Plotted again for visual clarity.

3.3.2 AMEND II

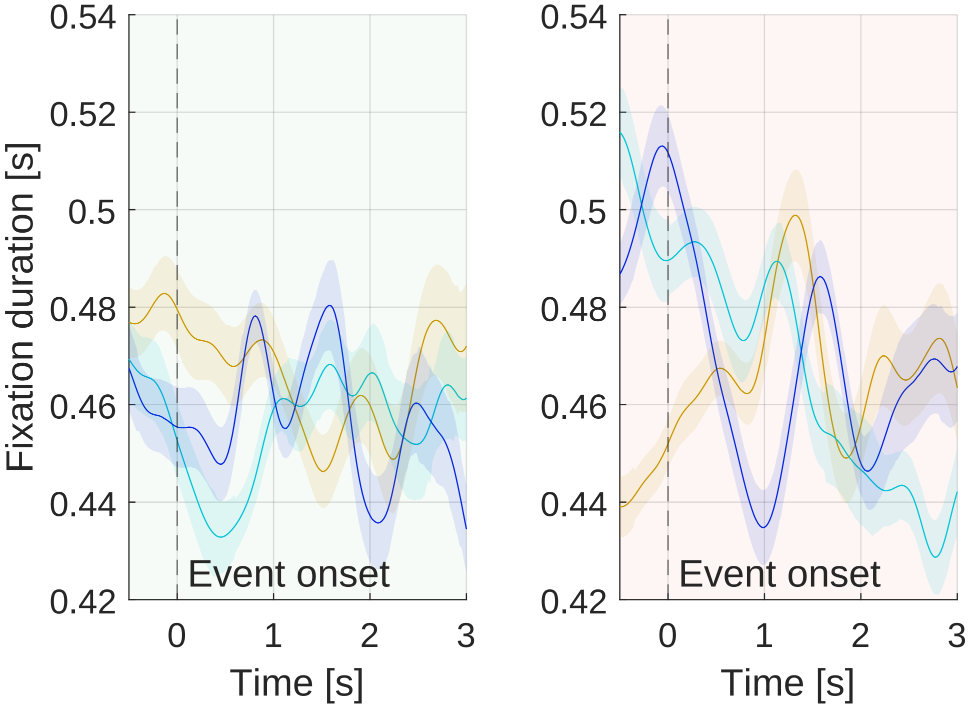


Figure 3.13: Fixation duration results from normal hearing (NH) TPs, averaged across participants separated by noise conditions, with events being either speaking or listening windows in AMEND II. Color legend is shown on fig. 3.7.

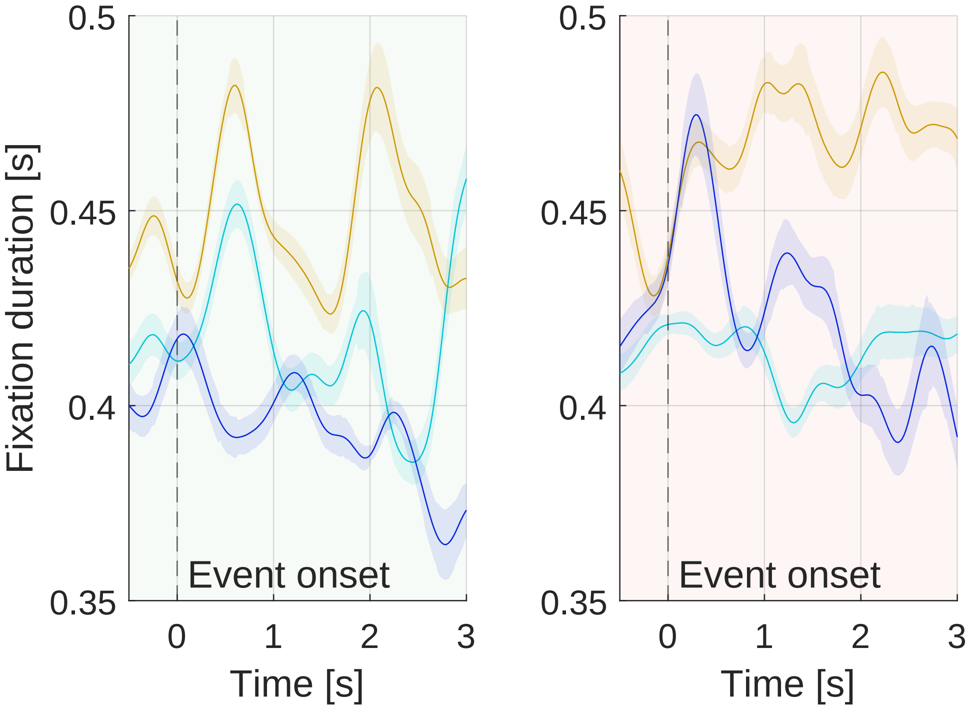


Figure 3.14: Fixation duration results from hearing impaired (HI) TPs, averaged across participants separated by noise conditions, with events being either speaking or listening windows in AMEND II. Color legend is shown on fig. 3.7.

4 Discussion

In this chapter, an examination of the findings presented in chapter 3 is conducted. With the main focus on testing the validity of the hypothesis initially proposed in section 1.3. The analysis of the results and their statistical measures were presented in order to give out a clear understanding of the research outcomes. This thesis can be summarized as an in-depth investigation into pupillometry and fixation measures, with a specific emphasis on the speaking and listening times for each participant. This characteristic is the primary difference with the analyses conducted within the AMEND study thus far.

Taking into consideration the absence of prior studies similar to AMEND, i.e., investigations involving physiological measures during large interactive communication periods between two unfamiliar talkers, and also, the distinctiveness of my thesis within the AMEND project, it can be confidently inferred that the pioneering outcomes presented hold considerable significance within the field. These findings will then provide a foundation in terms of expected results and response patterns if a similar study was to take place. Additionally, it is crucial to identify the factors that can clarify the observed behavioral outcomes, which could be related to the task, the subjects (and their hearing) or even the lab setup itself.

Upon analyzing the results from the fixation durations and pupil size measures together (as seen in section 3.3 and section 3.2 respectively), a significant synchronized negative correlation between these measures is evident for most of the results. This finding suggests that fixations and pupil size are potentially regulated by similar brain mechanisms, leading to synchronized behavioral responses. Additionally, it can also be stated that the findings provide substantial evidence to validate all of the hypothesis described in section 1.3, namely **H1**, **H2**, **H3** and **H4**. This holds true as most of the results align with the predicted outcomes and also demonstrate a consistent and statistically significant correlation. A more in-depth discussion on each measure will be presented later in section 4.1 and section 4.2.

4.1 Pupillometry

An article by Laeng et al. (2012) [36], based on some studies done with animals, suggested that tonic activity (non time-locked stimuli activity) of the Locus Coeruleus (LC), which is part of the brain that plays a major role in regulating arousal and attention [37], is associated with the tendency of switching to abandon a current task for another, while phasic activity is related to the processing of task-relevant events that are being attended. These activities can be indexed from pupil size measures and could explain most of the global results shown in section 3.2.

As for the AMEND I pupil size results, which are shown both in section 3.2.1 and section 3.2.3 respectively, it can clearly be seen that the most demanding noise conditions were ”N70” for speaking and ”SHL” for listening (as they elicited peaks in the responses) and also, in contrast, the least challenging condition was ”Quiet”. These results aligned with our initial expectations, thus providing a reference or benchmark for understanding the responses of NH participants in an interactive communication (for AMEND II). When speaking, increasing background noise exerts more effort, specially if it’s babble noise. This type of noise resembles speech from many different talkers which would potentially decrease intelligibility, demanding higher cognitive load. Moreover, it is likely that speakers will raise their voice in order to be understood by the listener. Listening becomes more demanding when the ear canal is obstructed by an earplug, eliciting a bigger pupil dilation than the challenge of the addition of background noise (without earplugs). This effect is likely attributed to the subjects' unfamiliarity with wearing earplugs, specially during conversations that require clear intelligibility.

AMEND II pupil diameter results, as shown in in section 3.2.2 and section 3.2.3, denote a large difference in cognitive load from NH and HI with the ”N70” noise condition when compared to ”Quiet” and ”N60”, both for speaking and listening times. This can clearly be seen in the non-baselined pupil diameter results shown in fig. 3.6a and fig. 3.8a. It is also significant to point out that there's a relatively similar increase in the pupil size features from one noise condition to the next (fig. 3.9), both for speaking and listening windows. Lastly, the impact of different hearing aid settings on the cognitive load of HI participants (as illustrated in fig. 3.10) is worth mentioning. Setting A demonstrates an improvement of variability in the features when compared to the un-aided condition, while setting B significantly reduces pupil diameter, indicating reduced effort which can be attributed to improved speech intelligibility. This contrast in results from one setting to another is explained with the fact that, setting A was designed to enhance the overall Signal-to-Noise Ratio (SNR) of the entire acoustic environment, while setting B prioritizes enhancing the SNR through directional focus towards the line of sight of the hearing aid user, this is a technique known as ”beam-forming”. The nature of the task itself, with both talkers always facing each other, will contribute to the positive effects of setting B, as beam-forming is most effective when the speech source is directly in front.

In the appendix, figures from fig. A.2 to fig. A.7 show the AMEND II pupil diameter responses over time, with and without baseline correction, from NH and HI participants combined with different hearing aids settings. These figures demonstrate a reduction in cognitive load among HI participants as they received increased noise reduction through their hearing aids. Furthermore, the figures reveal an adaptation process from NH participants based on whether the HI participants were wearing hearing aids (or not) and the specific settings used. Meaning that NH participants demonstrate a perceptible ability to recognize and adapt their communication strategies based on their own perception of the ease of communication of HI participants.

One of, if not the most important, topics of discussion in the processing of this thesis is the baseline. As stated in section 2.5.3, the baseline is defined as the 0.5 seconds time interval that precedes an event onset, that is, the start of a speaking or listening window. There are several reasons to support the selection of this baseline for the study. Firstly, previous evidence [21] states that, the baseline level from a TP steadily drops with time. In addition, having such large duration of trials, we realized that, there was a necessity for the baseline to differ not only between trials but also within a trial. Another reason for choosing this particular baseline is that, it was observed that most participants were able to predict the turns in the conversations, particularly when there is no overlap between or within events. Lastly, by choosing a baseline so closely related to events, we aim to capture and account for any potential anticipatory or preparatory processes that may occur. This ensures that the estimated baselined pupil size solely reflects the response from a subject to an event, that is, the difference between the preparatory phase leading up to the event and the event itself. One key limitation of this method is the potential overlap of baseline intervals with previous windows, meaning that, the baseline itself may be more associated with the preceding window rather than anticipating the upcoming one, resulting in distorted or even negative-peak baselined pupil responses.

4.2 Fixation duration

As seen previously on section 2.5.3, a novel study by Cui et al. (2023) [28] has managed to define a new measure of cognitive load by using eye gaze movements. In the study [28], the tasks were quite controlled in terms of fixation, stimuli and duration when comparing to the tasks in AMEND. The fixation duration findings from the study [28], suggest a correlation between fixations and cognitive load or listening effort, similar to the existing relationship also observed with pupil size. Likewise, the worse the noise conditions, the bigger the elicited response of fixation duration, which is exactly the same behavior as in pupil size. Previously, on section 4.1, we explored the challenges and limitations associated with establishing a reliable baseline, then, by employing fixation durations as a cognitive measure, we would observe responses that are independent of any baseline.

The analysis of the fixation duration results from both AMEND I and AMEND II (presented in section 3.3.1 and section 3.3.2, respectively), revealed that a majority didn't show a response that aligned with the anticipated results. The expected pattern, as demonstrated in the study by Cui et al. (2023) [28], suggested that higher task demands correspond to an increase in the fixation duration response. This finding may be closely linked to the nature of the task performed by the participants, which required them to intentionally focus on the diapix [14] pictures. The task demanded them to concentrate on the pictures either to comprehend any differences communicated by the other participant (listening effort) or to effectively describe any differences to the other participant (speaking effort).

Furthermore, it is important to note that the results, particularly for AMEND II, demonstrate fixation duration responses whose peaks are synchronized across different noise conditions. The consistency of this synchronization indicates that fixation duration is a reliable measure when it comes to portraying cognitive load. The results also indicate a distinction between speaking and listening windows, depicted by initial negative peaks and large initial positive peaks respectively. Additionally, the findings suggest that as the task demands increase, participants tend to spend less time fixating in speaking windows and, completely in opposite, participants fixate more in listening windows (in line with the results from Cui et al. (2023) [28]).

These observations suggest that participants may choose to avert their gaze from the picture while speaking, possibly to ensure a successful communication. In contrast, during listening, participants tend to fixate and scan the picture, which made it easier to compare the spoken information with the visual task, enabling them to distinguish differences faster. It was initially expected that lip reading would influence results during listening, especially as task demands increased. However, this assumption was proven false, with the exception of the HI participants (seen in fig. 3.14), which may already be accustomed to relying on lip reading on a daily basis in order to enhance speech understanding.

In the appendix, figures from fig. A.8 to fig. A.13 show the AMEND II fixation duration responses over time from NH and HI participants combined with different hearing aids settings. These figures further illustrate the previously mentioned adaptation process observed for NH participants in section 4.1. Additionally, the figures reveal a similarity in both shape and magnitude between the responses of HI participants without wearing hearing aids and those using setting B. This emphasizes how important settings are in specific environments, as they not only enhance frequencies affected with hearing loss but also enable HI participants to achieve a more ”natural” response by effectively suppressing unwanted noise. This phenomenon is similar to how the brain filters out unwanted stimuli through auditory processing mechanisms, particularly in challenging situations like the ”cocktail party problem” (described in section 1.1).

4.3 General observations

After addressing the particular aspects of the discussion related to pupillometry and fixation duration measures, both found in section 4.1 and section 4.2 respectively, we now take a broader perspective that will eventually lead to more general discussion points.

One examined aspect was the duration of any type of overlaps during conversations. As the pairs of participants did more trials, the percentage of overlaps decayed, indicating that naturalistic communication between participants involved minimal instances of simultaneous speech. This finding suggests that both NH and HI participants have developed effective strategies to cope with each other's communication styles, resulting in more refined and efficient communication dynamics over time. Another important aspect related to the interactive communication is the even distribution of utterances between talkers, assuring balanced conversations rather than the presence of a dominant ”leader” in the conversation. Meaning that, for most cases, individuals shared both participation to contribute and engagement in the dialogue.

The reason behind choosing the diapix [14] framework as the task, was to ensure an interactive communication environment during the trials, even if the conversations were solely related to the task. However, this introduces certain limitations in terms of the generalizability of the communication when compared to free-topic or spontaneous conversations, which are much similar to the natural conversations we aim to capture. Additionally, establishing a 4 minute limit for each trial aimed to minimize the effects of fatigue on participants, making sure that their performance and engagement remained consistent during trials.

To ensure a balanced and natural conversation, it was crucial to only recruit participants that were completely unfamiliar with each other. This inclusion criterion aimed to avoid a common ”shared language” among them, which is a phenomenon studied by Thomas et al. (2013) [38]. The aforementioned study suggests that shared languages are vital for enhancement and collaboration while communicating, which would contradict the intended purpose of capturing unbiased and authentic interactive conversations.

As mentioned earlier, this thesis serves as a pioneering project within the field, establishing a foundation for future work and research. Even thought there are various different aspects to polish, this thesis has provided a solid framework that can be utilized for similar studies in the future. However, we would like to propose suggestions for further exploration in future work within the field. Firstly, it would be interesting to extract completely different pupil size features, particularly those closely linked to response delays and condition-based variations. Additionally, improvements can be made when establishing the pupil baseline, by identifying specific regions of stability within the data. Moreover, the processing of speaking and listening windows could be changed for more precision and to add discrimination for non-relevant speech information. Lastly, we believe that some modifications to the test setup, including making the task more natural and changing all test conditions, could end up greatly enhancing the ecological validity of the study.

Lastly, it is important to note that in our analysis, we did not utilize the Bonferroni correction [39] to adjust the p-values for multiple comparisons. The Bonferroni correction [39] is a statistical method used to adjust p-values for multiple comparisons, reducing the risk of false-positive results. It increases the significance threshold for individual comparisons, which as stated in section 2.5.6, was of 0.05. However, due to the limited amount of individual subject data across conditions, the impact for not using the aforementioned correction on the statistical significance of our results, is unlikely to have affected the overall interpretation and statistical significance and robustness of the findings.

5 Conclusion

In this MSc thesis, a comprehensive analysis of pupil behavior and eye gaze movements during interactive communication has been conducted, complementing the ongoing AMEND research project [1]. The key finding of this research, since the beginning, has been the identification of behavioral or cognitive differences between speaking and listening in interactive communication. The analysis has concluded that these differences play a vital role in understanding the cognitive processes involved in communication. Additionally, the study has explored the distinction between arousal and attention, which greatly helped establishing a foundation for how nuanced mechanisms underlying communication dynamics work in natural conversations between two un-familiarized talkers.

To summarize, this master's thesis has effectively accomplished its primary research objectives. It presented and conducted a comprehensive analysis of results obtained from traditional pupil diameter measures and novel fixation duration measures. Holding promise for the development of improved treatment options and also laying a solid foundation for future investigations in the field.

Notably, the research revealed a strong correlation between cognitive load or effort, and environmental speech-shaped noise. As the noise conditions worsen, an increase is observed in both pupil and fixation behavioral measures. Research also revealed numerous behavioral differences in conversations between pairs of participants composed exclusively of NH individuals versus pairs consisting of one individual with HI and one NH individual. Meaning that, NH participants demonstrate distinct adaptive communication strategies when engaging in conversations with NH or HI participants.

The results from HI individuals wearing hearing aids demonstrate a decrease on cognitive load, which would potentially lead to an increase of intelligibility in noisy environments. This improvement is not only attributed to the use of hearing aids but also to selecting settings efficiently. Considering how large was the level in both noise conditions on the study, it was not surprising to find that setting B (AB), which had the greatest noise reduction, outperformed setting A (AA). Additionally, the same setting B also demonstrated a better performance in the quiet condition, which further proves the beneficial effect of hearing aids on any type of acoustic environment.

The performance of the chosen pupil size features, ranked in terms of ability to identify statistical different after fitting an LMM, were MPD, PPD and slope. However, it is noteworthy that certain features, such as MPD and PPD, appeared to convey more information when analyzing different test conditions. This suggests that some pupil size features are more effective than others, making them better suited for assessing and quantifying the cognitive demands associated with the task or condition being analyzed.

In conclusion, this MSc thesis contributes to the understanding of interactive communication by analyzing pupil behavior and eye gaze movements. The findings emphasize the differences between speaking and listening, as well as the relationship between arousal and attention. The results of this research provide valuable insights into behavioral communication effort and dynamics, as well as how to interpret distinct physiological measures under various test conditions and hearing aid settings. Looking ahead, future work can build upon these findings and explore within all aspects covered by this thesis, in order to achieve more naturalistic approaches to the study of interactive communication in real or natural conversational settings.

Bibliography

[1] Susan Aliakbary Hosseinabad (sulb@eriksholm.com) et al. *AMEND: Valid Outcome Measure for Communication Difficulty*. Eriksholm Research Centre, Linköping University (LiU), Technical University of Denmark (DTU), 2022.

[2] Ar Arnesen. “Presbyacusis: Loss of neurons in the human cochlear nuclei”. eng. In: *Journal of Laryngology and Otology* 96.6 (1982), pp. 503–511. ISSN: 17485460, 00222151. DOI: 10.1017/S002221510009277X.

[3] Gunay Kirkim et al. “Hearing loss and communication difficulty in the elderly”. eng. In: *Mediterranean Journal of Otology* 3.3 (2007), pp. 126–132. ISSN: 13055267.

[4] Stuart R. Steinhauer and Gad Hakerem. “The Pupillary Response in Cognitive Psychophysiology and Schizophrenia”. eng. In: *Annals of the New York Academy of Sciences* 658.1 Psychophysiol (1992), pp. 182–204. ISSN: 17496632, 00778923. DOI: 10.1111/j.1749-6632.1992.tb22845.x.

[5] Tepring Piquado, Derek Isaacowitz, and Arthur Wingfield. “Pupillometry as a measure of cognitive effort in younger and older adults”. In: *Psychophysiology* 47 (3 2010), pp. 560–569. ISSN: 14698986. DOI: 10.1111/j.1469-8986.2009.00947.x.

[6] Avashna Govender and Simon King. “Using pupillometry to measure the cognitive load of synthetic speech”. In: vol. 2018-September. International Speech Communication Association, 2018, pp. 2838–2842. DOI: 10.21437/Interspeech.2018-1174.

[7] Soroosh Solhjoo et al. “Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load”. In: *Scientific Reports* 9 (1 Dec. 2019). ISSN: 20452322. DOI: 10.1038/s41598-019-50280-3.

[8] M. P. Paulraj et al. “Auditory evoked potential response and hearing loss: A review”. eng. In: *Open Biomedical Engineering Journal* 9 (2015), pp. 17–24. ISSN: 18741207. DOI: 10.2174/1874120701509010017.

[9] World Health Organization (WHO). *Deafness and hearing loss*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (visited on 06/19/2023).

[10] Eckhard H. Hess and James M. Polt. “Pupil Size as Related to Interest Value of Visual Stimuli”. In: *Science* 132.3423 (1960), pp. 349–350. ISSN: 00368075, 10959203. URL: http://www.jstor.org/stable/1706082 (visited on 02/20/2023).

[11] Eckhard H. Hess and James M. Polt. “Pupil Size in Relation to Mental Activity during Simple Problem-Solving”. In: *Science* 143.3611 (1964), pp. 1190–1192. ISSN: 00368075, 10959203. URL: http://www.jstor.org/stable/1712692 (visited on 02/20/2023).

[12] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. “A Simplest Systematics for the Organization of Turn-Taking for Conversation”. und. In: *Language* 50.4 (1974), p. 696. ISSN: 15350665, 00978507. DOI: 10.2307/412243.

[13] Stephanie Haro et al. “EEG Alpha Power and Pupil Diameter Reflect Endogenous Auditory Attention Switching and Listening Effort”. In: (2021). DOI: 10.1101/2021.07.29.453646. URL: https://doi.org/10.1101/2021.07.29.453646.

[14] Rachel Baker and Valerie Hazan. “DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs”. eng. In: *Behavior Research Methods* 43.3 (2011), pp. 761–770. ISSN: 15543528, 1554351x. DOI: 10.3758/s13428-011-0075-y.

[15] MATLAB R2021a. *version: 9.10.0.2198249 (R2021a) - Update 8*. Natick, Massachusetts: The MathWorks Inc., 2021.

[16] Ziad S. Nasreddine et al. “The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment”. eng. In: *Journal of the American Geriatrics Society* 53.4 (2005), pp. 695–699. ISSN: 00028614, 15325415. DOI: 10.1111/j.1532-5415.2005.53221.x.

[17] Judith Jaeger. “Digit symbol substitution test”. eng. In: *Journal of Clinical Psychopharmacology* 38.5 (2018), pp. 513–519. ISSN: 1533712x, 02710749. DOI: 10.1097/JCP.0000000000000941.

[18] Sven Hilbert et al. “The digit span backwards task: Verbal and visual cognitive strategies in working memory assessment”. eng. In: *European Journal of Psychological Assessment* 31.3 (2015), pp. 174–180. ISSN: 21512426, 10155759. DOI: 10.1027/1015-5759/a000223.

[19] Jens Bo Nielsen and Torsten Dau. “The Danish hearing in noise test”. eng. In: *International Journal of Audiology* 50.3 (2010), pp. 202–208. ISSN: 17088186, 14992027. DOI: 10.3109/14992027.2010.524254.

[20] Raul Sanchez-Lopez et al. “Towards Auditory Profile-Based Hearing-Aid Fittings: BEAR Rationale and Clinical Implementation”. eng. In: *Audiology Research* 12.5 (2022). Ed. by Agnieszka Szczepek, pp. 564–573. ISSN: 20394349, 20394330. DOI: 10.3390/audiolres12050055.

[21] Matthew B. Winn et al. “Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started”. eng. In: *Trends in Hearing* 22 (2018), pp. 1–22. ISSN: 23312165. DOI: 10.1177/2331216518800869.

[22] Mariska E. Kret and Elio E. Sjak-Shie. “Preprocessing pupil size data: Guidelines and code”. eng. In: *Behavior Research Methods* 51.3 (2019), pp. 1336–1342. ISSN: 15543528, 1554351x. DOI: 10.3758/s13428-018-1075-y.

[23] Christophe Leys et al. “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median”. eng. In: *Journal of Experimental Social Psychology* 49.4 (2013), pp. 764–766. ISSN: 10960465, 00221031. DOI: 10.1016/j.jesp.2013.03.013.

[24] Helia Relaño-Iborra and Per Bækgaard. “PUPILS pipeline: A flexible Matlab toolbox for eyetracking and pupillometry data processing [arXiv]”. eng. In: *Arxiv* (2020), 4 pp.

[25] Kenneth Holmqvist et al. “Eye tracking: empirical foundations for a minimal reporting guideline”. mul. In: (2023).

[26] Jeff Klingner, Barbara Tversky, and Pat Hanrahan. “Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks”. In: *Psychophysiology* 48 (3 2011), pp. 323–332. ISSN: 14698986. DOI: 10.1111/j.1469-8986.2010.01069.x.

[27] Jackson Beatty and Brennis Lucero-Wagoner. “The pupillary system.” In: (2000).

[28] M. Eric Cui and Björn Herrmann. “Eye movements decrease during effortful speech listening”. In: *bioRxiv* (2023). DOI: 10.1101/2023.02.08.527708. eprint: https://www.biorxiv.org/content/early/2023/02/09/2023.02.08.527708.full.pdf. URL: https://www.biorxiv.org/content/early/2023/02/09/2023.02.08.527708.

[29] RP OSHEA. “THUMB RULE TESTED - VISUAL ANGLE OF THUMBS WIDTH IS ABOUT 2 DEG”. eng. In: *Perception* 20.3 (1991), pp. 415–418. ISSN: 14684233, 03010066. DOI: 10.1068/p200415.

[30] A. Josefine Munch Sørensen, Michal Fereczkowski, and Ewen N. MacDonald. “Effects of Noise and Second Language on Conversational Dynamics in Task Dialogue”. eng. In: *Trends in Hearing* 25 (2021), p. 23312165211024482. ISSN: 23312165. DOI: 10.1177/23312165211024482.

[31] Stephen C. Levinson and Francisco Torreira. “Timing in turn-taking and its implications for processing models of language”. eng. In: *Frontiers in Psychology* 6 (2015), p. 731. ISSN: 16641078. DOI: 10.3389/fpsyg.2015.00731.

[32] Emanuel A. Schegloff. “Overlapping talk and the organization of turn-taking for conversation”. eng. In: *Language in Society* 29.1 (2000), pp. 1–63. ISSN: 00474045, 14698013. DOI: 10.1017/s0047404500001019.

[33] Agustín Gravano, Julia Hirschberg, and Štefan Běnuš. “Affirmative cue words in task-oriented dialogue”. eng. In: *Computational Linguistics* 38.1 (2012), pp. 1–39. ISSN: 15309312, 08912017. DOI: 10.1162/COLI_a_00083.

[34] Muneeb Imtiaz Ahmad et al. “A framework to estimate cognitive load using physiological data”. eng. In: *Personal and Ubiquitous Computing* (2020), pp. 1–15. ISSN: 16174917, 16174909. DOI: 10.1007/s00779-020-01455-7.

[35] Henrik Singmann and David Kellen. “An Introduction to Mixed Models for Experimental Psychology”. und. In: *New Methods in Cognitive Psychology* (2019). DOI: 10.4324/9780429318405-2.

[36] Bruno Laeng, Sylvain Sirois, and Gustaf Gredebäck. “Pupillometry: A window to the preconscious?” eng. In: *Perspectives on Psychological Science* 7.1 (2012), pp. 18–27. ISSN: 17456924, 17456916. DOI: 10.1177/1745691611427305.

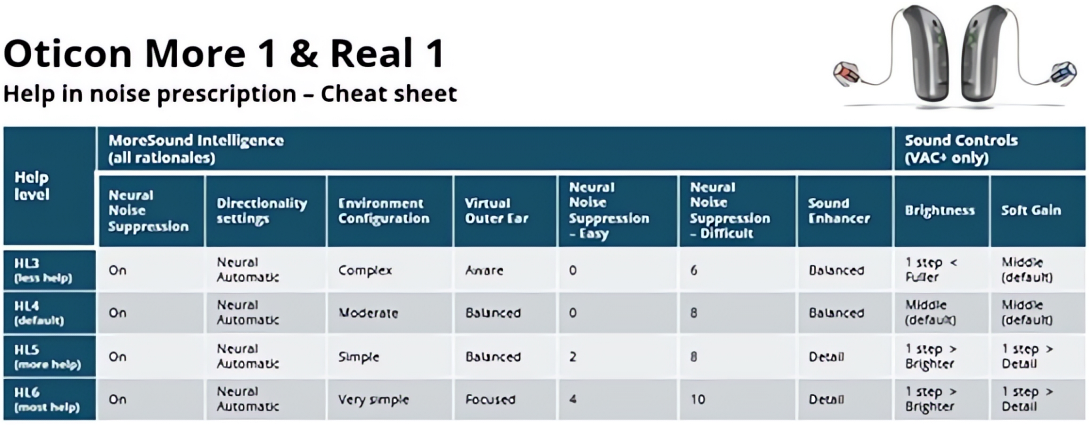
[37] Jennifer A. Ross and Elisabeth J. Van Bockstaele. “The Locus Coeruleus- Norepinephrine System in Stress and Arousal: Unraveling Historical, Current, and Future Perspectives”. eng. In: *Frontiers in Psychiatry* 11 (2020), p. 601519. ISSN: 16640640. DOI: 10.3389/fpsyt.2020.601519.

[38] Joyce Thomas and Deana McDonagh. “Shared language:Towards more effective communication”. eng. In: *Australasian Medical Journal* 6.1 (2013), pp. 46–54. ISSN: 19361935, 18361935. DOI: 10.4066/AMJ.2013.1596.

[39] C.E. Bonferroni. *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze. Seeber, 1936. URL: https://books.google.dk/books?id=3CY-HQAACAAJ.

A Appendix

Table A.1: Different Help levels in noise prescription with both NAL-NL2 and VAC+.



Oticon More 1 & Real 1

Help in noise prescription – Cheat sheet

Help level	MoreSound Intelligence (all rationales)							Sound Controls (VAC+ only)	
	Neural Noise Suppression	Directionality settings	Environment Configuration	Virtual Outer Ear	Neural Noise Suppression – Easy	Neural Noise Suppression – Difficult	Sound Enhancer	Brightness	Soft Gain
HL3 (less help)	On	Neural Automatic	Complex	Aware	0	6	Balanced	1 step < Fuller	Middle (default)
HL4 (default)	On	Neural Automatic	Moderate	Balanced	0	8	Balanced	Middle (default)	Middle (default)
HL5 (more help)	On	Neural Automatic	Simple	Balanced	2	8	Detail	1 step > Brighter	1 step > Detail
HL6 (most help)	On	Neural Automatic	Very simple	Focused	4	10	Detail	1 step > Brighter	1 step > Detail

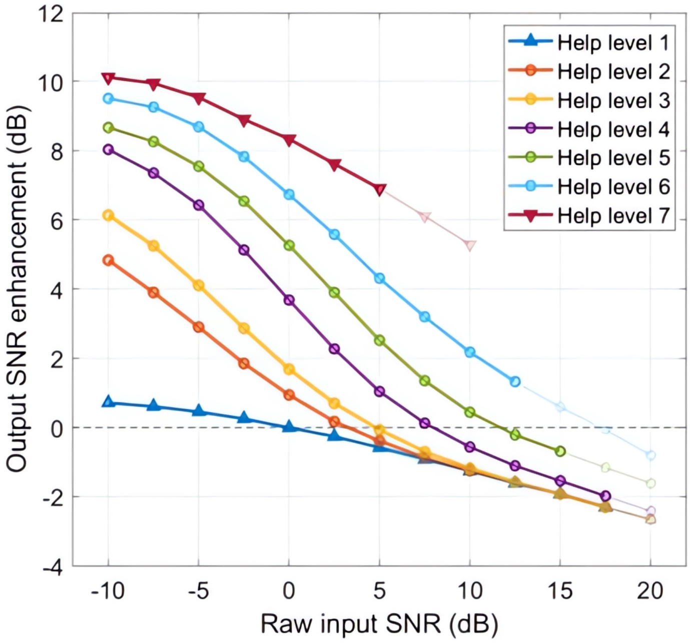
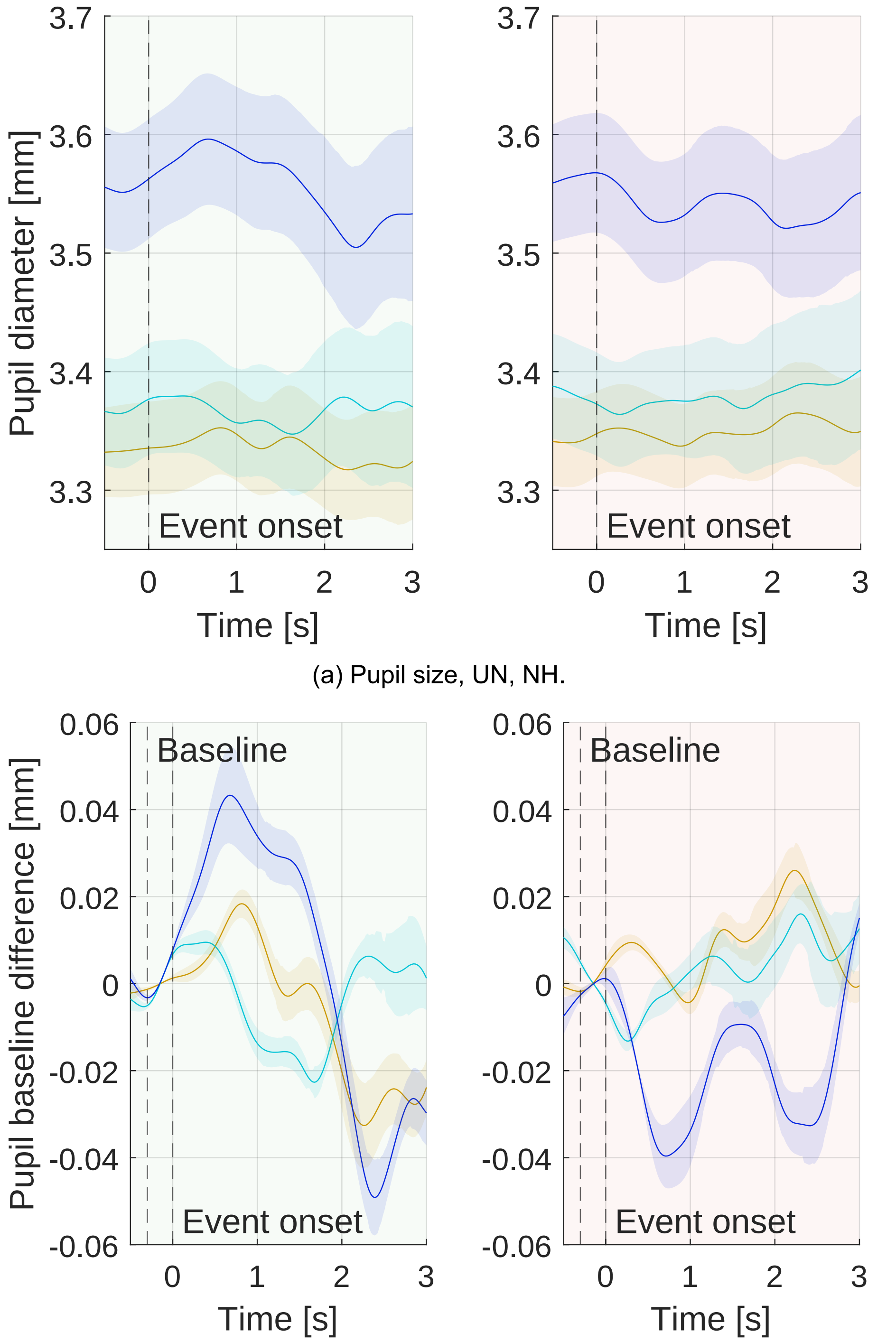


Figure A.1: Output SNR enhancement as a function of input SNR for each HL [20].

A.0.1 AMEND II - Additional results

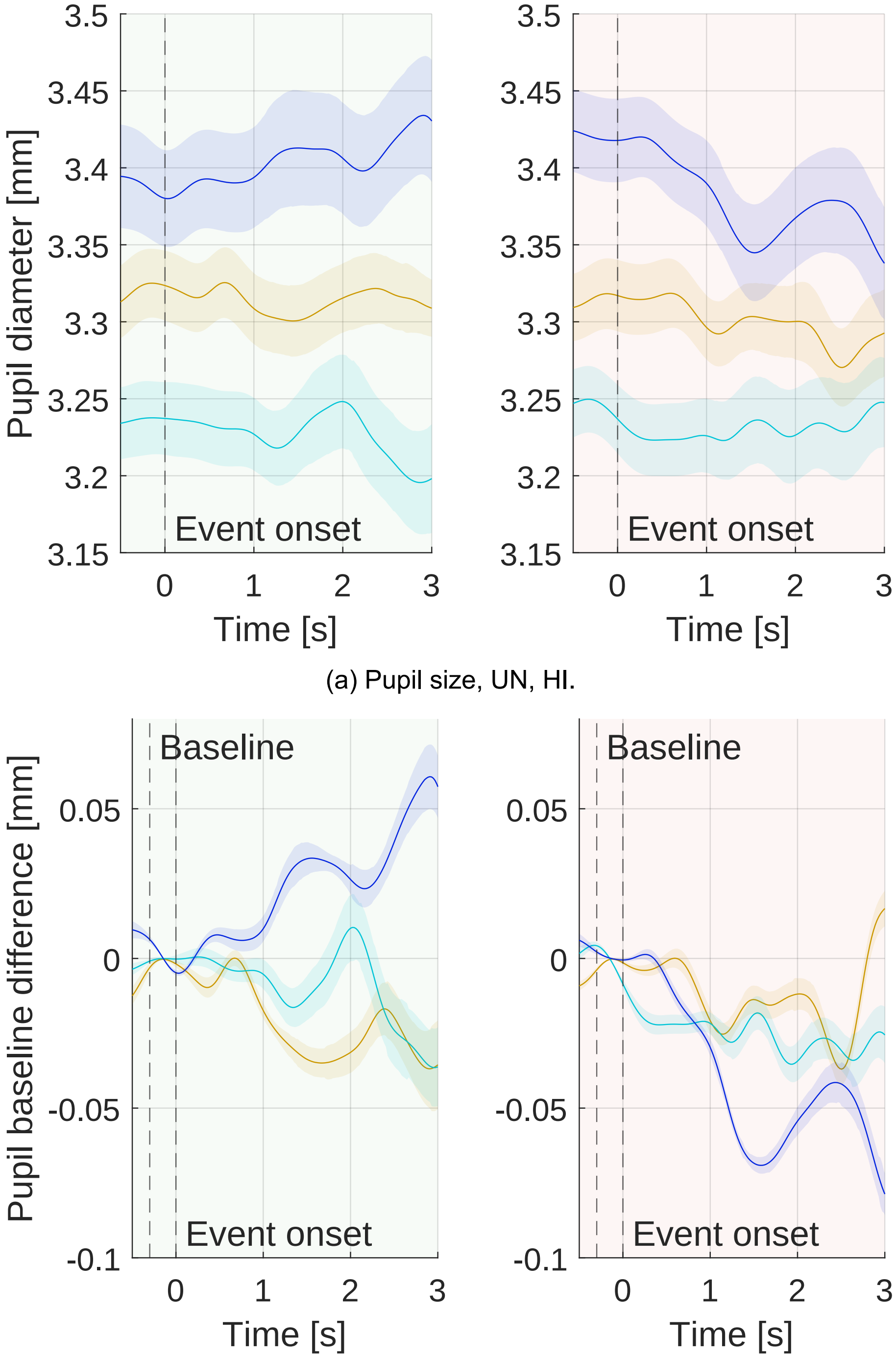
Pupil diameter



(a) Pupil size, UN, NH.

(b) Baselined pupil size, UN, NH.

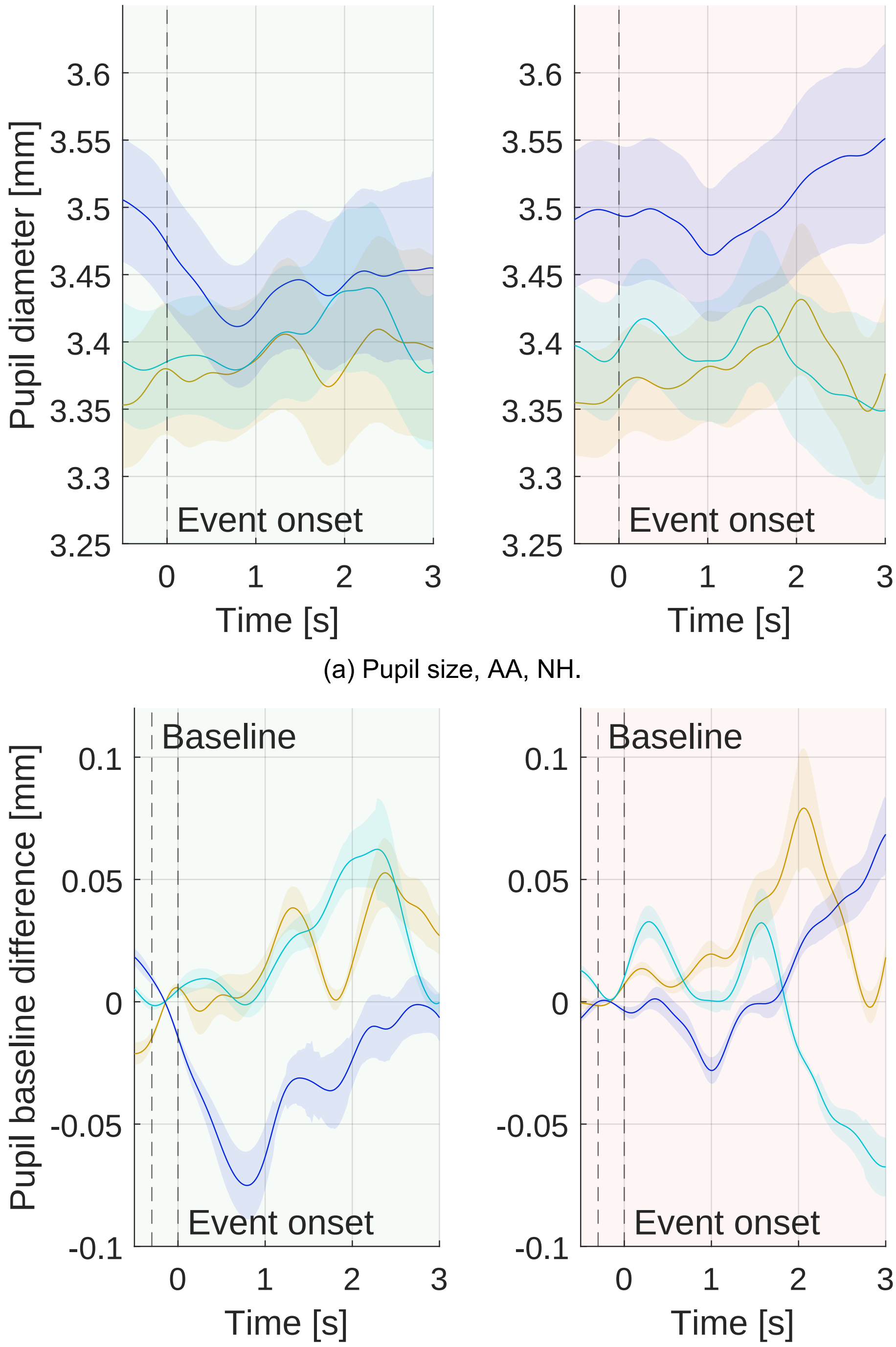
Figure A.2: Normal hearing (NH), in un-aided condition (UN), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.



(a) Pupil size, UN, HI.

(b) Baselined pupil size, UN, HI.

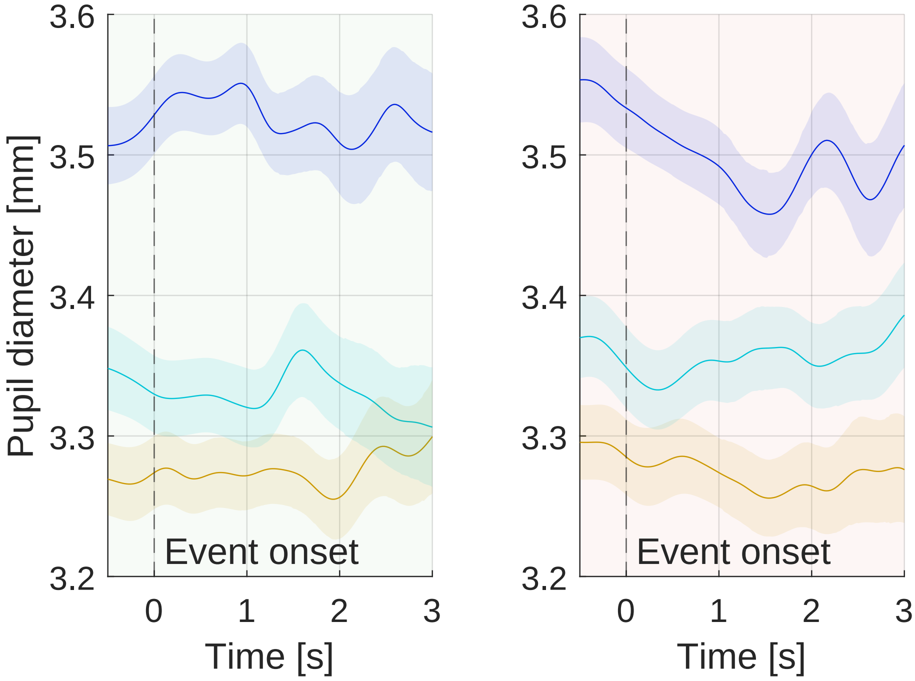
Figure A.3: Hearing impaired (HI), in un-aided condition (UN), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.



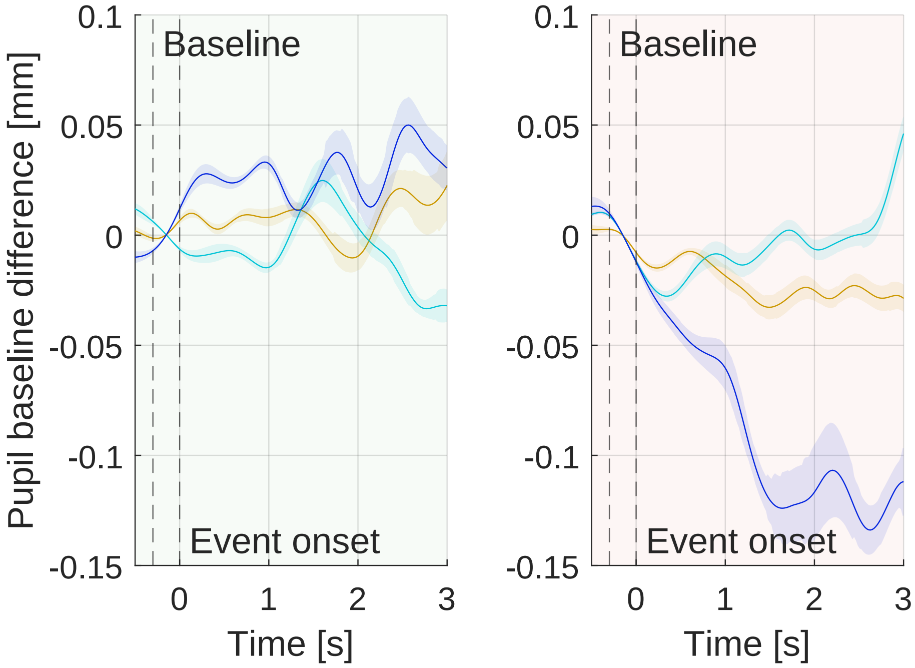
(a) Pupil size, AA, NH.

(b) Baselined pupil size, AA, NH.

Figure A.4: Normal hearing (NH), in aided setting A condition (AA), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

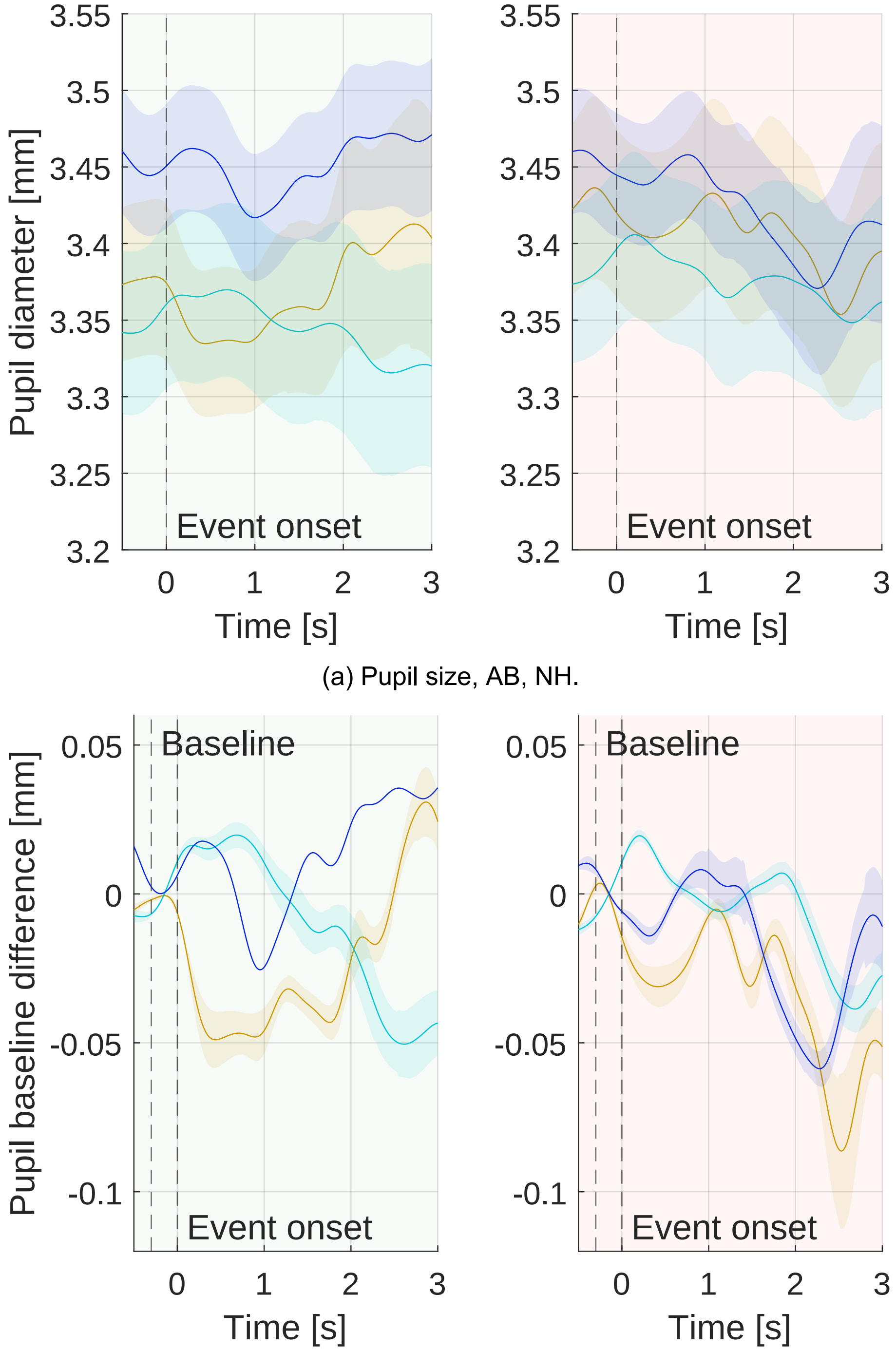


(a) Pupil size, AA, HI.



(b) Baselined pupil size, AA, HI.

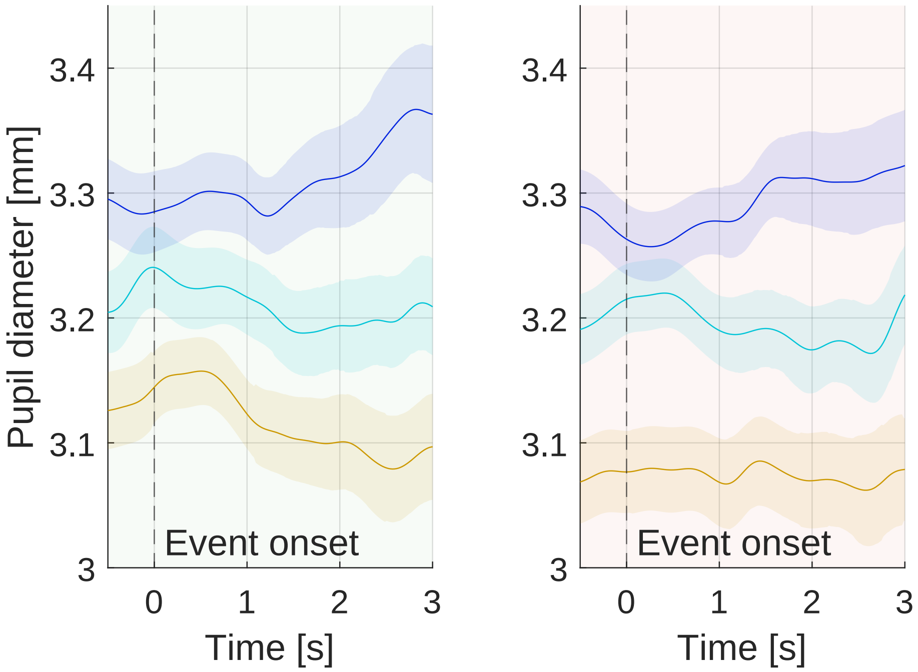
Figure A.5: Hearing impaired (HI), in aided setting A condition (AA), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.



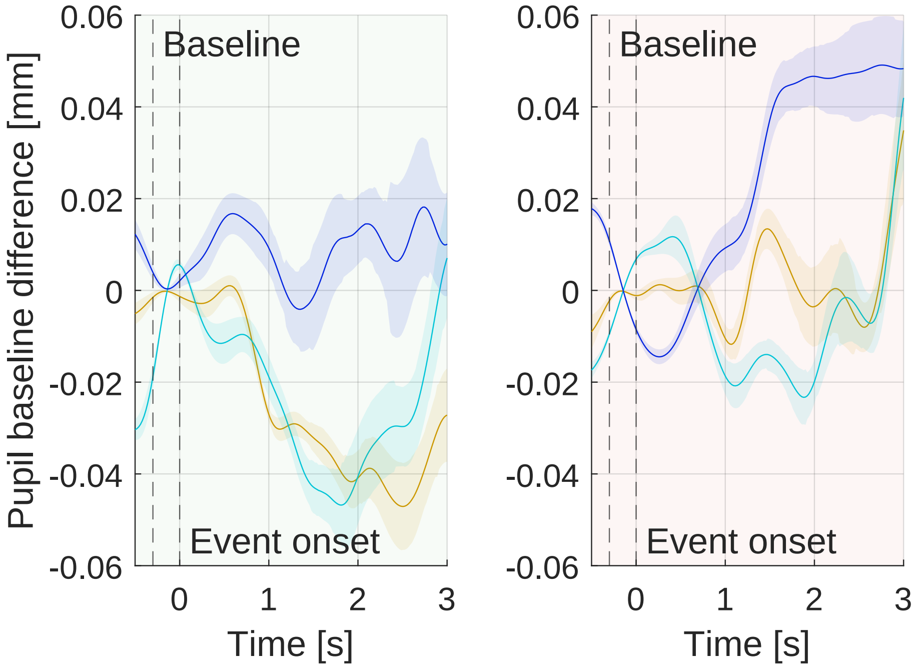
(a) Pupil size, AB, NH.

(b) Baselined pupil size, AB, NH.

Figure A.6: Normal hearing (NH), in aided setting B condition (AB), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.



(a) Pupil size, AB, HI.



(b) Baselined pupil size, AB, HI.

Figure A.7: Hearing impaired (HI), in aided setting B condition (AB), pupil size results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

Fixation duration

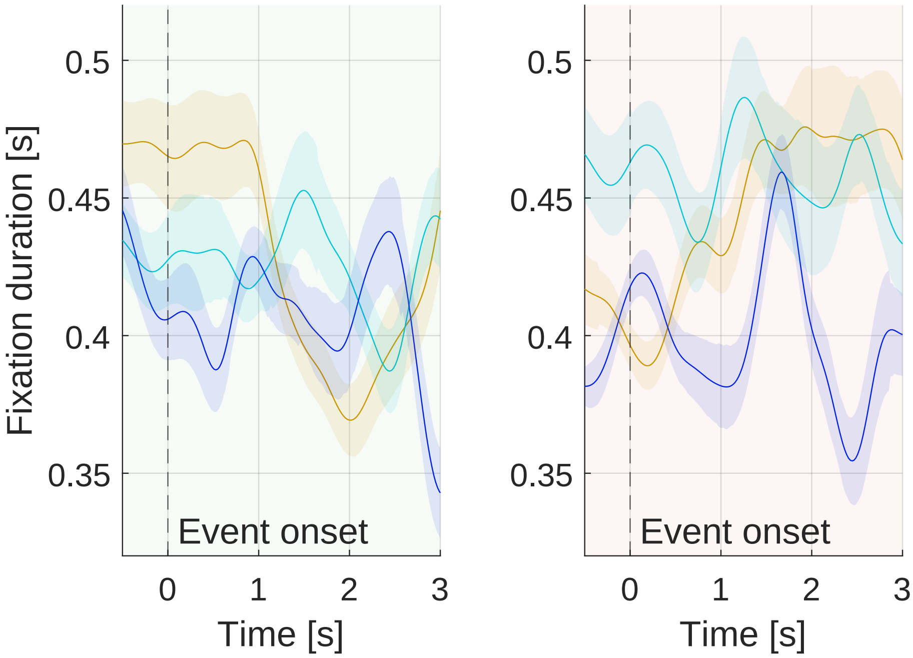


Figure A.8: Normal hearing (NH), in un-aided condition (UN), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

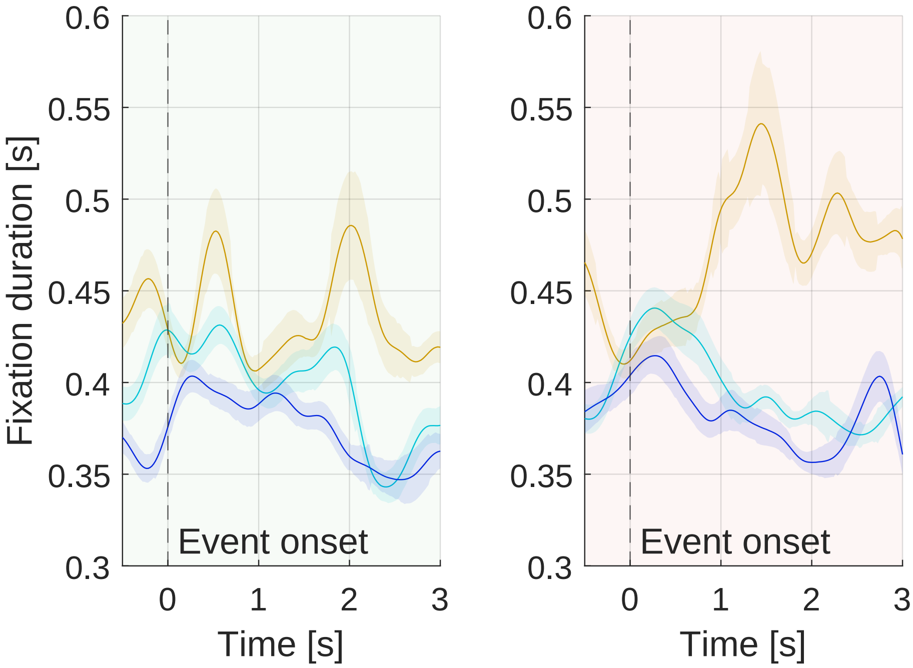


Figure A.9: Hearing impaired (HI), in un-aided condition (UN), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

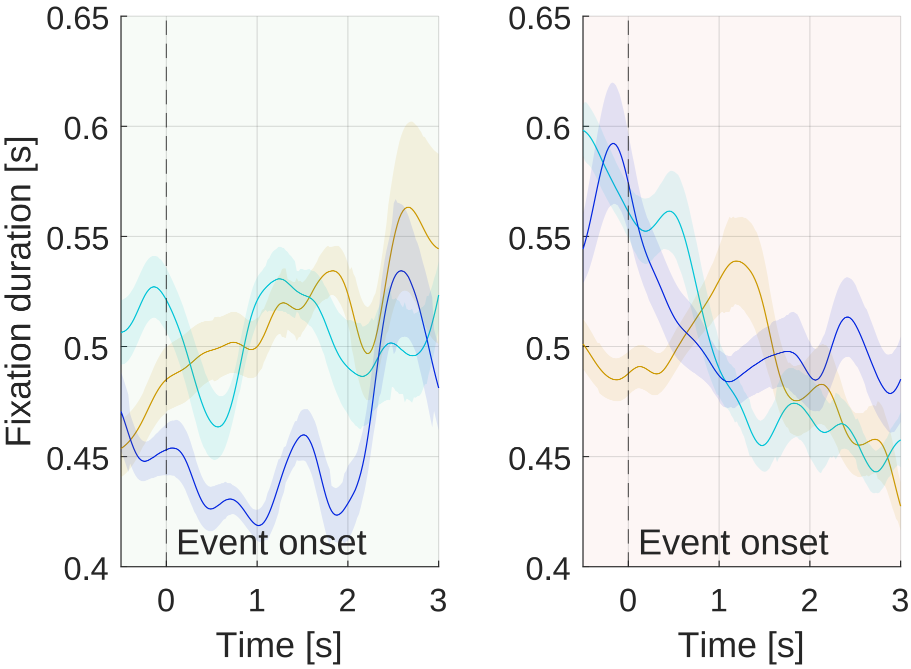


Figure A.10: Normal hearing (NH), in aided setting A condition (AA), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

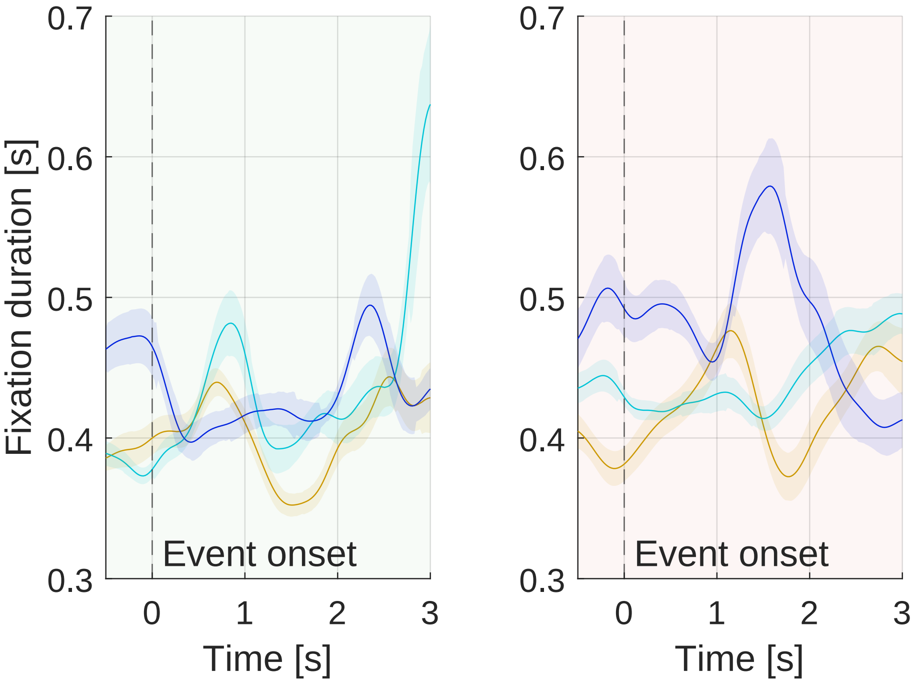


Figure A.11: Hearing impaired (HI), in aided setting A condition (AA), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

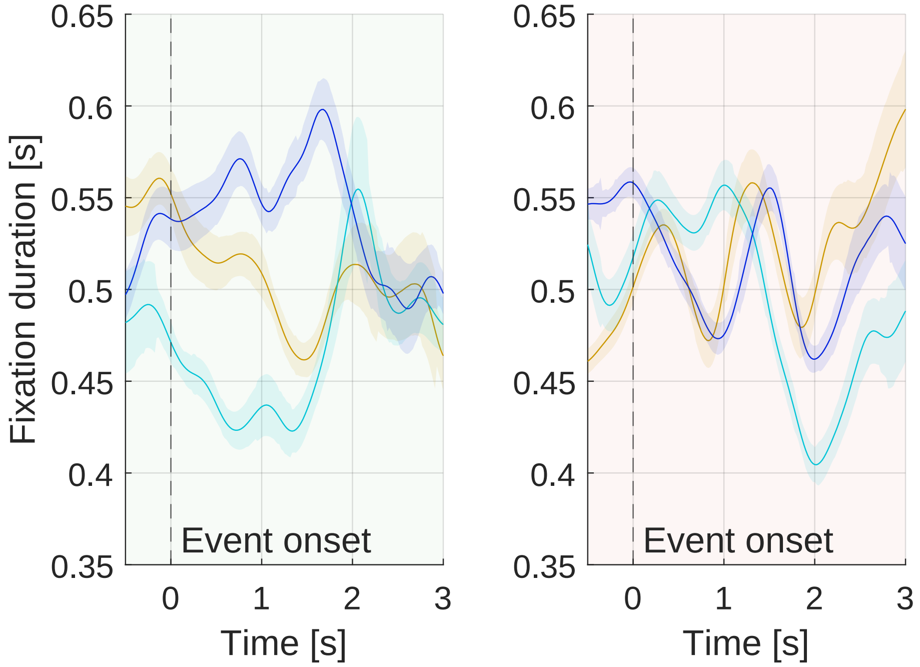


Figure A.12: Normal hearing (NH), in aided setting B condition (AB), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

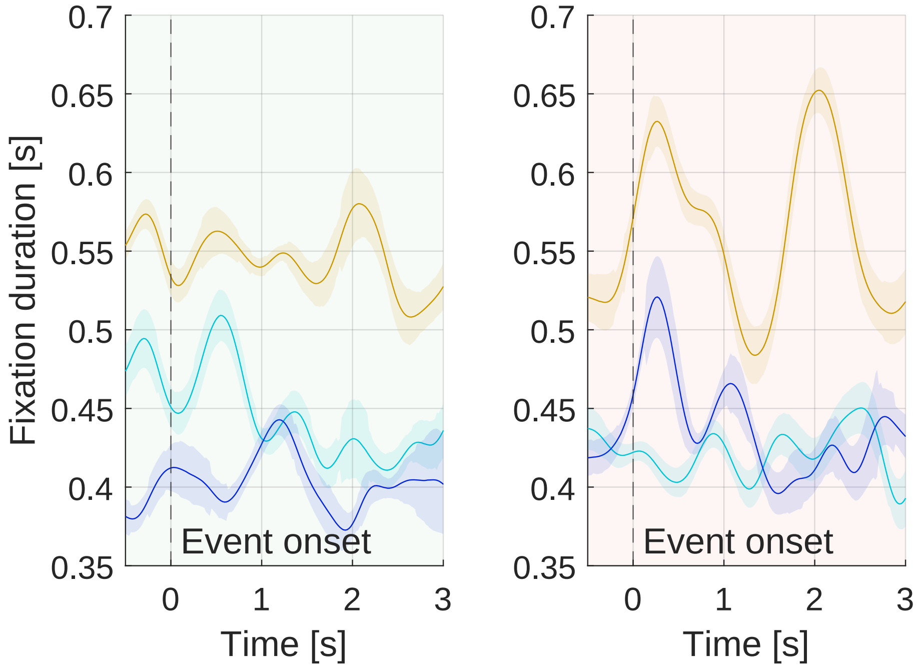


Figure A.13: Hearing impaired (HI), in aided setting B condition (AB), fixation duration results averaged across participants and separated by noise conditions, with events being either speaking or listening windows (left and right panels, respectively) in AMEND II. Color legend is shown on fig. 3.7.

Technical
University of
Denmark

Ørsteds Plads, Building 352
2800 Kgs. Lyngby
Tlf. 4525 1700

www.hea.healthtech.dtu.dk